A PUBLICATION OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION

THE UNIVERSITY OF MICHIGAN
NOV 1 1961
UNDERGRADUATE LIBRARY

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

UNDERGRADUATE LIBRA

ARTICLES

The Generality of Immunity to Persuasion Produced by Pre-exposure to Weakened Counterarguments 475
Demetrios Papageorgis and William J. McGuire
Incidental versus Focal Presentation of Drive Related Stimuli 482
Fred Pine
The Relationship between Word Frequency and Perceptibility with a Forced-Choice Technique 491
Janet A. Taylor, Doris C. Rosenfeldt, and Rudolph W. Schulz
Preconscious Perception: A Re-examination of the Poetzl Phenomenon 497
Harold Johnson and Charles W. Eriksen
Stability of Achievement and Recognition Seeking Behaviors from Early Childhood through Adulthood 504
Howard A. Moss and Jerome Kagan
A Reinterpretation of Some Pathological Disturbances in Conceptual Breadth 514
Loren J. Chapman
The Generality of the WAIS Standardization for the Aged: A Regional Comparison 520
Carl Eisdorfer and Louis D. Cohen
Interpersonal Perceptions of Repressors and Sensitizers and Component Analysis of Assumed Dissimilarity Scores 528
John Altrocchi
Dominance as a Factor in Achievement and Perception in Cooperative Problem Solving Interactions 535
William T. Smelser
Delay of Gratification, Need for Achievement, and Acquiescence in Another Culture 543
Walter Mischel
Interference Effects with Schizophrenic Subjects in the Acquisition and Retention of Verbal Material 553
John W. Donahoe, Mary Ellen Curtin, and Leonard Lipton

Continued on inside front cover

This is the last issue of Volume 62.

Title Page and Contents for the Volume appear herein.

Volume 62 May 1961 Number 3

CONTENTS—*Continued*

Disagreement with Others, Their Credibility, and Experienced Stress 559
Harold B. Gerard

Time Perspective, Consistency of Attitude, and Social Influence 565
Harold B. Gerard and George S. Rotter

Familiarity and the Order of Presentation of Persuasive Communications 573
Robert E. Lana

Opinion Evaluation and Affiliation 578
Roland Radloff

Fear and Social Comparison 586
Harold B. Gerard and Jacob M. Rabbie

A Comparison of Generalization and Adaptation-Level as Theories of Connotation 593
Jerome E. Podell

Psychotherapeutic and Other Hospital Treatment Criteria: The Dilemma 598
Ralph P. Forsyth and George W. Fairweather

Intrajudge Consistency in Ratings of Heterogeneous Persons 605
Cecily F. Gross

An Experimental Test of Schema Interaction 611
Donald P. Spence

Attitudes toward Old People in an Older Sample 616
Nathan Kogan

A Cross-Linguistic Investigation of Phonetic Symbolism 623
Murray S. Miron

Pancultural Factor Analysis of Reported Socialization Practices 631
Leigh Minturn Triandis and William W. Lambert

Sources of Frustration and Targets of Aggression: A Cross-Cultural Study 640
Leigh Minturn Triandis and William W. Lambert

The Perpetuation of an Arbitrary Tradition through Several Generations of a Laboratory Microculture 649
Robert C. Jacobs and Donald T. Campbell

Differences in Developmental Experience, Defensiveness, and Personality Organization between Two Classes of Problem Drinkers 659
Richard R. Rudie and Laurence S. McGaughran

The Assessment of Anxiety in Psychological Experimentation: A Theoretical Reformulation and Test 666
George C. Rosenwald

Dimensions of Individual Performance in Small Groups under Task and Social-Emotional Conditions 674
Richard D. Mann

CRITIQUE AND NOTES

Sex, Ethnicity, and Play Preferences of Preschool Children 683
Boyd R. McCandless and June M. Hoyt

A Comparison of Mental Retardates and Normals on Kinesthetic Figural Aftereffects 686
Herman H. Spitz and Ronald S. Lipman

Operant Conditioning of Verbal Behavior of Dysthymics and Hysterics 688
Donna M. Gelfand and C. L. Winder

Perceptual Selectivity, Memory, and Anti-Semitism 690
Lee Pulos and Bernard Spilka

Perceived Characteristics of Elected Representatives 693
Robert R. Blake and Jane Srygley Mouton

The Interaction of Personality and Muscle Tension in Producing Agreement on Commonality of Verbal Associations 696
James L. Kuethe

A Factor Analytic Investigation of Personality Variables and Reminiscence in Motor Learning. 698
A. W. Bendig and Charles R. Eigenbrode

New Zealand Children's Attitudes towards Maoris 701
G. M. Vaughan and R. H. T. Thompson

Perceived Differences in Task Competence and Asymmetry of Dyadic Influence 705
Melvyn D. Croner and Richard H. Willis

Successive Responses to Simultaneous Stimulation in Elderly Patients with Memory Disorder. 709
James Inglis and R. E. Sanderson

Interpersonal Attraction and Attitude Similarity 713
Donn Byrne

Anxiety, Birth Order, and Susceptibility to Social Influence 716
Frederick R. Staples and Richard H. Walters

The Influence of Valence on Distance Judgments of Human Faces 720
John Beloff and Halla Beloff

... 559
... 565
... 573
... 578
... 586
... 593
... 598
... 605
... 611
... 616
... 623
... 631
... 640
ory
... 649
be-
... 659
on
... 666
nal
... 674
... 683
... 686
... 688
... 690
... 693
ty
... 696
g. 698
... 701
... 705
er. 709
... 713
... 716
... 720

Volume 62 May 1961 Number 3

Journal of Abnormal and Social Psychology

Editor: M. Brewster Smith, *University of California, Berkeley*
Associate Editor: E. J. Shoben, Jr., *Teachers College, Columbia University*

Consulting Editors

Gordon W. Allport
Harvard University

Urie Bronfenbrenner
Cornell University

Roger Brown
Massachusetts Institute of Technology

Donald T. Campbell
Northwestern University

Launor F. Carter
System Development Corporation

Richard Christie
Columbia University

Morton Deutsch
Bell Telephone Laboratories

I. E. Farber
State University of Iowa

Donald W. Fiske
University of Chicago

N. L. Gage
University of Illinois

Roger W. Heyns
University of Michigan

Robert R. Holt
New York University

Irving L. Janis
Yale University

Harold H. Kelley
University of Minnesota

Leonard S. Kogan
Community Service Society of New York

Richard S. Lazarus
University of California, Berkeley

Daniel J. Levinson
Harvard Medical School

David C. McClelland
Harvard University

Leo Postman
University of California, Berkeley

Eliot H. Rodnick
University of California, Los Angeles

Nevitt Sanford
University of California, Berkeley

Seymour Sarason
Yale University

Richard L. Solomon
University of Pennsylvania

published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $4.00.

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Department of Psychology
University of Michigan
Ann Arbor, Michigan

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 475-481

THE GENERALITY OF IMMUNITY TO PERSUASION PRODUCED BY PRE-EXPOSURE TO WEAKENED COUNTERARGUMENTS[1]

DEMETRIOS PAPAGEORGIS
University of Illinois

AND WILLIAM J. McGUIRE
Department of Social Psychology, Columbia University

THE assumption that people tend to defend their beliefs by avoiding exposure to contrary points of view has been widely accepted (Festinger, 1957; Hovland, 1954; Janis, 1957; Klapper, 1949, 1957–58). Experimental support for this view is not as impressive as one might wish, considering the extent that it has been relied on for theoretical purposes, but several studies—summarized by Festinger (1957, Ch. 6, 7) or inspired by his dissonance theory (Brodbeck, 1956; Cohen, Brehm, & Latané, 1959; Ehrlich, Guttman, Schönbach, & Mills, 1957; Mills, Aronson, & Robinson, 1959)—have provided some experimental corroboration. Since a person's practice of defending his beliefs by avoidance should tend to leave him unmotivated and inexperienced in developing supporting arguments or in refuting counterarguments, it is not surprising to find that while beliefs prove highly vulnerable in situations of forced exposure, they tend to remain relatively unaffected by persuasion campaigns that allow self-selected exposure (Hovland, 1959).

Besides suggesting a reason for the vulnerability of beliefs, the assumption of selective exposure has implications regarding the relative effectiveness of various techniques for immunizing beliefs in advance against forced exposure to massive doses of counterarguments. Resistance to such counterarguments would be analogous to the resistance to disease of a person brought up in an aseptic environment. It has been found (Lumsdaine & Janis, 1953; McGuire & Papageorgis, 1961), in keeping with this analogy, that a person's belief is more effectively immunized against a later massive dose of counterarguments by a resistance-stimulating pre-exposure to the counterarguments, weakened by explicit refutation, than by prior familiarization with arguments supporting the belief.

The earlier study by the present authors demonstrated heightened resistance to subsequent strong forms of the same counterarguments as those to which the person had been pre-exposed. The present study tests the hypothesis that such pre-exposure to weakened counterarguments produces resistance to subsequent strong forms of alternative counterarguments as well. There are at least two reasons for predicting the generalized immunization. First, the experience of having seen some counterarguments against the belief effectively refuted may lower the impressiveness even of alternative counterarguments. Second, a person's pre-exposure experience may make him more aware that his belief is indeed vulnerable to attack and therefore motivate him to develop supporting arguments that make his belief more resistant even to alternative counterarguments presented later. A secondary purpose of the study was to test the validity of each of these two conjectured mechanisms of the hypothesized generalized immunity effect.

METHOD

General Procedure

The experiment was presented to the subjects as an attempt to obtain standardization data for a newly developed test of scientific information and aptitude with emphasis on verbal skills. Each subject took part in two half-hour experimental sessions one week apart. The first session was designed to expose the subject to an immunizing treatment defending a cultural truism, and the second, to strong counterarguments attacking this (or another) truism.

First session. In the first, or "immunizing," session each subject was required to read an immunizing message refuting a pair of counterarguments against a truism. The subject was allowed 6 minutes in which to read the 500–550-word refutational message and to select and underline the crucial clause in each paragraph. Our main interest was implied to be in how the subject performed at this selecting and underlining task. At the end of the first session, the subject was asked to complete an attitude questionnaire measuring 22 beliefs including the experimentally crucial ones. The attitude scales are described more fully below. The belief scores obtained on the immunized issue during this first session

[1] This research was supported by a grant from the National Science Foundation, Office of Social Sciences.

furnished a measure of the direct effect of the "immunizing" treatment itself, prior to the strong counterargument message which was presented in the second session. It was stressed that the subject was to indicate his own belief and not necessarily what the essay assigned to him might have argued. Subjects were told that these opinions were being measured because an incidental purpose of the study was to determine if a person's skill in reading comprehension was affected by his beliefs regarding the material that he read. Upon completion of this session, the subject was asked to refrain from any discussion of the test and to return the following week.

Second session. One week later, the subject took part in the second session which included the presentation of the message containing strong counterarguments against his belief. To test the main experimental hypothesis (that the prior refutation of counterarguments tends to immunize the beliefs not only against the specific counterarguments refuted, but also against other counterarguments directed at the same belief) some subjects received strong versions of the same counterarguments that had been refuted in the previous session, while other subjects received strong forms of previously unmentioned alternative counterarguments. The subject was again allowed 6 minutes in which to read strong counterarguing messages (again between 500 and 550 words in length) and to select and underline the crucial clause in each paragraph, his performance on this underlining task once more being represented as our main interest. He then filled out again the opinion questionnaire used in the first session and was told to indicate his present personal belief, regardless of what had been argued in any of the passages that he had read. The beliefs that the subject indicated on the second administration provided the measure of the relative effectiveness of the strong counterarguments on immunized and nonimmunized beliefs.

After this task, the subject was asked to list all the arguments that came to mind in favor of the critical truisms. In addition, he was asked to rate the strong counterargument message he had just read in the second session with respect to several evaluative dimensions (the validity of its conclusion, the amount of evidence presented, etc.). The purpose of these two tasks was to test the conjectured mechanisms of the generalized immunity effect.

Finally, the subject was asked to answer five questions designed to determine the extent to which he had guessed the true purpose of the experiment, had discussed the experiment with others, and felt that the time allowed for the various tasks had been adequate.

At the end of the experiment, considerable effort was made to inform the subject of the nature of the experiment and of the deceits used. Particular stress was put on the fact that both immunization and counterargument messages were constructed for propaganda purposes and that no weight should be given to any argument simply because it had been included in one of these propaganda messages.

Materials

Issues. Two issues dealing with public health were selected (on the basis of data from a prestudy) for extremeness and homogeneity of precommunication beliefs within the student population being sampled. The beliefs selected were the following: "Everyone should get a chest X ray each year in order to detect any possible tuberculosis symptoms at an early stage" and "Everyone should brush his teeth after every meal if at all possible."

The subjects indicated their beliefs about these statements by marking an "X" in the appropriate space on a 15-interval graphic scale labeled periodically from "Definitely false" on the extreme left to "Definitely true" on the extreme right. In computing the results reported below, the intervals were scored 1, 2, . . . , 14, 15, going from left (false) to right (true). The preliminary data showed that the mean belief on each of the two issues was over 13, with the mode at 15 on the 15-point scale.

Messages. Immunizing and strong counterargument messages[2] were prepared on each belief. The immunizing messages consisted of three paragraphs: the first mentioning the two "misleading" counterarguments against the belief and the second and third containing refutations of these two counterarguments. Two immunizing messages were prepared on each belief, each refuting a different pair of counterarguments, so that for some subjects the belief could be subsequently attacked by strong forms of the previously refuted counterarguments and for other subjects, by strong forms of new counterarguments.

The strong counterargument messages were similar in structure to the immunizing messages. Each contained three paragraphs, the first of which briefly described two counterarguments against the belief, and the following two paragraphs strongly elaborated these counterarguments with specific evidence for their validity. Here again to implement the experimental design, there were two strong counterargument messages on each issue, each message elaborating an alternative pair of counterarguments. All the messages, both immunizing and strong counterarguing, were between 500 and 550 words in length.

Design

There was a total of 16 experimental conditions representing all possible combinations of the four immunizing messages in the first session and the four strong counterarguing messages in the second. All these combinations were required to test the effect of some experimental treatment, to counterbalance material, or to obtain control data. Since some of the conditions entered into more frequent or more crucial comparisons than others, the subjects were not distributed in equal numbers over the conditions (see Table 1). In addition to conditions required for counterbalancing the materials, the design provided final opinion scores after each of five different hypothesis-relevant treatments.

[2] The complete text of the eight communications used in this study has been deposited with the American Documentation Institute. Order Document No. 6769 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

immunization, then presentation of the strong forms of the same counterarguments; immunization, then presentation of different strong counterarguments; immunization only; strong counterarguments only; and neither immunization nor counterarguments.

Subjects

A total of 73 subjects, all students enrolled in various undergraduate courses in the summer session at the University of Illinois, took part in both sessions of the experiment. Data from 9 additional subjects who took part in the first (immunization) session but did not appear for the second (strong counterargument) session are omitted from the analysis.

RESULTS AND DISCUSSION

Adequacy of the Experimental Conditions

Post-experimental questionnaire results. An attempt was made to disguise the inductions of opinion change as a test of verbal skills. In response to a final question concerning what else the test might have been used to measure, 16 of the 73 subjects gave responses indicating that they might possibly have perceived that the experiment involved persuasion. Actually, fewer subjects may have been suspicious at the time the opinions were being measured, since between the opinion measure and this final question there intervened the requirements of listing favorable arguments and rating the quality of the strong counterargument message, tasks that may well have increased the perception of persuasive intent. At any rate, the data from these 16 "suspicious" subjects show no reliable differences from those obtained from the remaining 57 subjects.

The responses on the post-experimental questionnaire also indicated good compliance with the instructions not to discuss the test between sessions. Only 8 subjects admitted any discussion and this was largely of an irrelevant type such as telling others how long the session lasted. The subjects also indicated that the time was adequate for all parts of the test, except for the section requiring the listing of all the supporting arguments they could think of about the truisms; 10 subjects complained that the allowed time for this section (5 minutes) was inadequate.

Adequacy of the opinion items. The control data indicate that the initial beliefs on the two critical issues were, as intended, quite extreme in the sample. The first session mean belief score in the no immunization condition was 13.29 on the 15-point scale for the two crucial issues. The submeans for the tooth brushing and chest X ray issues were, respectively, 14.03 and 12.53, and the modal belief was 15 for each issue. In the control condition where neither immunization nor strong counterarguments were communicated on a given issue, there was a slight decline in belief from first to second session, the 13.29 mean of the first session declining to 13.23 in the second, a fluctuation that falls far short of conventionally accepted levels of statistical significance.

To measure the reliability of the type of opinion scale used in this study, one opinion item was repeated twice on different pages of the questionnaire. This item dealt with the contagiousness of mental illness and was selected as the reliability check because a previous study (McGuire & Papageorgis, 1961) found that it behaved very similarly to the two crucial issues used in this study. Since the check item appeared twice in each session, each subject responded to it four times. The two means on the 15-point scale in the first session were 13.51 and 13.67; in the second, 13.51 and 13.63, none of which differences remotely approach conventional levels of statistical significance. The test-retest coefficient of correlation in the first session was 0.81 and in the second 0.89. The higher reliability in the second administration ($.10 > p > .05$, based on a test of difference between z transformations) may reflect a practice effect that reduced random errors in responding.

Effectiveness of the strong counterarguments. The two beliefs, despite their high initial acceptance, proved quite vulnerable to the strong counterarguments in the second session. Where strong counterarguments were not preceded by first session immunization, the beliefs were reduced to a mean of 5.73 (see Table 1), representing a drop of 7.50 points from the 13.23 second session mean in the control group where neither immunization nor strong counterarguments were given ($p < .001$). As pointed out earlier, this high vulnerability of the beliefs in the absence of preparatory treatments can be explained by assuming that the subject normally minimized cognitive conflict by avoiding exposure of his beliefs to counterarguments, and thus is unprepared to deal with them when exposure is enforced by the experimental conditions.

TABLE 1

BELIEFS BEFORE AND AFTER EXPOSURE TO IMMUNIZING AND/OR STRONG COUNTERARGUMENT MESSAGES
(Numbers in the cells show mean belief strength on a 15-point scale, 15 indicating strong adherence to belief)

Session	Strong Counterarguments Only—No Prior Immunization	Immunizing Pre-exposure to Refuted Counterarguments, Followed by Exposure to Strong Forms of the:		Immunizing Pre-exposure Only—No Later Exposure to Strong Forms of the Counterarguments	Control—No Immunization and No Later Strong Counterarguments
		Same Counterarguments	Alternative Counterarguments		
First	—	—	—	(a) 14.34 (73)[a]	(b) 13.29 (73)
Second	(c) 5.73 (26)	(d) 9.25 (20)	(e) 8.70 (27)	(f) 14.15 (26)	(g) 13.23 (47)

Note.—Significance levels of theoretically relevant comparisons (probability values shown are all based on analysis of variance and a two-tailed test): a vs. b, $p = .02$; a vs. f, *ns*; b vs. g, *ns*; c vs. d, $p < .01$; c vs. e, $p < .05$; c vs. f, $p < .001$; c vs. g, $p < .001$; d vs. e, *ns*; d vs. g, $p < .001$; e vs. g, $p < .001$.

[a] Numbers in parentheses give number of cases on which the cell mean is based.

Effects of the Prior Defensive Treatments

Direct effect. The immediate effect of the refutational message (prior to the presentation of the strong counterarguing message) has particular interest because of the possibility of a "boomerang" effect. A priori, such a defense might actually weaken the belief since it involves acquainting the subject with some novel counterarguments (albeit refuted) against a belief he had regarded as a truism, beyond attack. The results indicate that the direct effect of the refutational defense is, on the contrary, to strengthen the belief: the beliefs at the end of the first session averaged 14.34 points in the refutational defense condition and 13.29 points in the control conditions that involved no immunization (Table 1). This direct gain of 1.05 points, small in absolute magnitude, is impressive when we consider the low "ceiling" under which it occurred. A gain of this magnitude represents a change of over 61% of the remaining scale distance and is significant beyond the .02 level. (Significance levels in this report are based on two-tailed analysis of variance tests.) Hence fears that pre-exposure to the refuted counterarguments might boomerang seem unfounded.

Immunizing effects. Aside from this direct strengthening effect, the refutational defense was found to increase considerably the person's resistance to the subsequent strong counterargument messages. The overall effect of the strong counterarguments in the prior immunization conditions was to reduce the beliefs to 8.94 points, representing a drop of 4.29 points ($p < .001$) below the belief level in the no immunization, no counterargument control conditions. Prior immunization, while far from completely effective, left the beliefs 3.21 points higher ($p < .01$) than the 5.73 point level to which they were reduced when the strong counterargument had not been preceded by prior defense. The immunization procedures, therefore, attenuated the effect of the strong counterarguments by 43%. The direction of these differences, and even their absolute magnitudes, follow closely those found in the previous McGuire and Papageorgis (1961) study.

To test the main hypothesis that a generalized immunization effect was produced by the pre-exposure, some subjects were presented in the second session with strong forms of the same counterarguments that had been refuted in the first session, while other subjects received strong forms of alternative counterarguments against the belief. The resistance developed against the same counterarguments was slightly greater than that against the alternative counterarguments, the beliefs after both immunization and strong counterarguments being 9.25 and 8.70, respectively, in the two conditions (see Table 1). The small size of this difference (which does not approach the conventional levels of statistical significance) indicates that the general resistance developed by the pre-exposure is only slightly, if at all, less than the specific immunity. The 8.70 level to which the beliefs were reduced in the alternative-counterargument immunization conditions was significantly greater ($.02 < p < .05$) than the 5.73 level to which thes

counterarguments reduced the beliefs in the no immunization condition. In view of the triviality of this difference between resistance to same and to alternate counterarguments, there would seem to be little necessity, in practice, to guess beforehand what counterargument might be used at a future time against a belief we wish to immunize: the prior refutation of some counterarguments develops almost as much resistance to subsequent strong forms of alternative counterarguments as to the same counterarguments.

A second test of the generalized immunity hypothesis is provided by additional data collected in the experiment. Besides the general opinion question, there were four additional questions on each issue, each of the four measuring belief in one of the four counterarguments. Hence at the end of both sessions it was possible to measure not only the subject's belief on the general issue, but his specific beliefs on the four counterarguments—the two he heard refuted as well as the two that had not been previously refuted. In conditions where the subject received neither immunization nor strong counterarguments on an issue, the mean tendency to disagree with the specific counterarguments was 10.62 on the 15-point scale. When the subject received the strong counterarguments without any previous immunization on the given issue, the mean tendency to disagree with the counterargument was reduced to 6.41 ($p < .001$). When the strong counterarguments came after the same pair of counterarguments that had been refuted, the mean tendency to disagree was 8.46; and where the strong counterarguments had been preceded by refutation of the alternative pair of counterarguments this mean was 8.60, both of which means are significantly greater ($p < .01$) than the 6.41 level in the no immunization conditions. Hence the prior experience of hearing some counterarguments refuted actually produced (trivially) more resistance—8.60 as compared to 8.46—to subsequent strong forms of the alternative counterarguments than to the very same counterarguments. These data on the agreement with the specific counterarguments corroborate the findings on the main issues as regards a highly generalized immunization.

The generalizability of these results regarding immunization against persuasion is limited by several considerations. Immunization has been demonstrated only with respect to specific preselected beliefs and not across issues. Also, the persistence of the immunity conferred by pre-exposure to refuted counterarguments remains to be investigated: the present study demonstrates only that appreciable immunization remains in effect after one week. Furthermore, the results of this study were obtained on culturally accepted truisms, issues of a type for which previous defense by avoidance of exposure to counterarguments seems most probable. It is possible also that the results of this study, obtained on college students, may not be applicable to less "select" samples. Questions might also be raised regarding the independence of the two sets of counterarguments in interpreting the finding of generalized immunity.

Hypothesized Mechanism for the Generalized Immunity Effect

Counterargument ratings. One hypothesized mechanism of the generalized immunity effect was that pre-exposure during the first (immunization) session to impressive refutations of some counterarguments would cause the subject to be less impressed by any subsequent counterarguments against the belief. To test whether the strong counterarguments presented in the second session did indeed seem less credible to subjects who had the experience of witnessing refutations of alternative counterarguments against the same belief, additional ratings were obtained from the subjects. Each subject was asked to rate on a seven-point scale the quality and credibility of the counterargument message presented in the second session. There were 10 questions touching on different aspects of the counterarguments so that the most favorable possible score was 70 and the least favorable 7. Without immunization, the mean rating was 53.69 points; with immunization, 45.96 points. The difference between these two means is significant beyond the .01 level. In the immunization conditions the ratings of the alternate (not previously refuted) counterarguments were lowered to 46.52, which is almost as low as the 45.20 rating of the very counterarguments that had been refuted. (The difference

between these two means is not significant.) The ratings of the strong counterarguments became significantly lower ($p < .05$) than the 53.69 rating in the no prior immunization condition. Hence one way that immunization "generalizes" is by causing strong forms of counterarguments against a belief to appear less impressive to the subject after he has seen refutations of some counterarguments against the same belief, whether these strong counterarguments are the same or different from those previously refuted.

Number and quality of favorable arguments. A further mechanism hypothesized to account for the generalized immunity effect predicted that pre-exposure to the counterarguments in refuted forms would make the subject realize that his "truism" could be attacked and thus stimulate him to think up more arguments in support of his belief. At the end of the second session the subjects were asked to list arguments favoring the initial belief on each issue. With immunization, the mean number of arguments listed was 2.62, whereas without immunization the mean dropped to 2.32. This difference, although in the predicted direction, did not attain the conventional levels of statistical significance ($.10 < p < .20$). The arguments listed by the subjects were also rated as to their quality on a 0–5 scale by an independent rater, a high school teacher unfamiliar with the experiment. With immunization, the mean quality rating was 3.19 whereas in the controls it dropped to 3.01, another difference which though not significant is in the predicted direction. The lack of significance may be partially due to a failure to allow the subjects sufficient time for listing the arguments. Since the specific time complaints made in the post-experimental questionnaire were directed exclusively against this part of the "test," the lack of time may have allowed an excessive contribution to the total variance of individual difference variables unrelated to the experimental manipulation, such as intelligence, speed of writing, etc.

The results of the experiment, therefore, give clear indication of a generalized immunity conferred by pre-exposure to some refuted counterarguments against the given belief. They also indicate that one mechanism of the generalized resistance is the tendency of pre-exposure to reduce the subsequent impressiveness even of new counterarguments against the belief. Not ruled out as an additional mechanism is stimulation by pre-exposure of the development of supporting arguments.

SUMMARY

A previous study showed that strong initial beliefs are more effectively immunized against persuasion by pre-exposing them to counterarguments weakened by refutation than by pre-exposure to supporting arguments. The present study tested the hypothesis that pre-exposure to refutations of some counterarguments against the belief would have a generalized immunization effect, making the beliefs more resistant to strong doses not only of the specific counterarguments refuted but also of alternative counterarguments against the given belief. Such a generalized effect was hypothesized to be the result of lowering the perceived credibility of later attacks against the belief and stimulating the person to think up supporting arguments for his belief.

The 73 college student subjects participated in two experimental sessions, one devoted to pre-exposure to refuted counterarguments, and a second, one week later, to the presentation of strong counterarguments. The design allowed subjects to be exposed to counterarguments only, immunization only, or immunization followed either by the identical or alternate counterarguments.

As expected, the beliefs proved highly vulnerable to the strong counterarguments when there was no prior immunization. Immunization had a direct strengthening effect on the beliefs and also substantially reduced the effect of the subsequent strong counterarguments. Pre-exposure to refutations of some counterarguments produced significant resistance to later strong forms of, not only the same counterarguments, but alternative counterarguments as well, in accord with the predicted effect of generalized immunity. Regarding the hypothesized mechanisms of such generalized immunity, the quality and credibility of the strong counterarguments were rated lower by subjects who had received immunization than by those who had not, even where the strong counterarguments were different from those previously refuted. There was also a tendency

short of statistical significance, for immunized subjects to be able to think up more supporting arguments, and ones of higher quality, for their beliefs than nonimmunized subjects.

REFERENCES

BRODBECK, MAY. The role of small groups in mediating the effects of propaganda. *J. abnorm. soc. Psychol.*, 1956, **52**, 166–170.

COHEN, A. R., BREHM, J., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *J. Pers.*, 1959, **27**, 63–73.

EHRLICH, D., GUTTMAN, I., SCHÖNBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, **54**, 98–102.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

HOVLAND, C. I. Effects of the mass media of communication. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 2. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 1062–1103.

HOVLAND, C. I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *Amer. Psychologist*, 1959, **14**, 8–17.

JANIS, I. L. Motivational effects of different sequential arrangements of conflicting arguments: A theoretical analysis. In C. I. Hovland (Ed.), *Order of presentation in persuasion*. New Haven: Yale, 1957. Pp. 170–186.

KLAPPER, J. Effects of the mass media. New York: Columbia University, Bureau of Applied Social Research, 1949. (Mimeo)

KLAPPER, J. What we know about the effects of mass communication: The brink of hope. *Publ. opin. Quart.*, 1957–58, **21**, 453–474.

LUMSDAINE, A. A., & JANIS, I. L. Resistance to "counterpropaganda" produced by a one-sided versus a two-sided "propaganda" presentation. *Publ. opin. Quart.*, 1953, **17**, 311–318.

MCGUIRE, W. J., & PAPAGEORGIS, D. The relative efficacy of various types of prior belief-defense in producing immunity to persuasion. *J. abnorm. soc. Psychol.*, 1961, **62**, 327-337.

MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *J. abnorm. soc. Psychol.*, 1959, **59**, 250–253.

(Received June 20, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 482–490

INCIDENTAL VERSUS FOCAL PRESENTATION OF DRIVE RELATED STIMULI[1]

FRED PINE

Research Center for Mental Health, New York University

THE contribution made by consciousness to processes of stimulus intake and subsequent thought is of central importance to dynamic theories of thinking—witness the roles of conscious and nonconscious processes in creativity (cf. Kris, 1952) and the role of incidental stimuli in dream formation (Freud, 1955). Much current research has demonstrated that stimuli with various marginal relationships to consciousness have effects on thinking processes. These effects have been shown in studies of subliminal stimuli (Eagle, 1959; Paul & Fisher, 1959; Smith, Spence, & Klein, 1959), of partially seen stimuli (Fisher, 1957; Shevrin & Luborsky, 1958), and of incidental stimuli (Pine, 1960). Klein (1959) has extensively explored the implications of these studies for a theory of consciousness.

Effects shown in studies of marginal stimuli have often been distinctively different from the more veridical effects of consciously received and adaptively relevant stimuli. The present study deals with the *contrasting* effects of incidental and focal stimuli upon subsequent thought processes. It follows from an earlier one (Pine, 1960) in which incidental stimuli appeared in subsequently produced stories in transformed ways while focal stimuli appeared in more direct ways. The stimulus passages used had been written to emphasize symbolic qualities: a description of a cow emphasizing its oral-passive aspect and description of a hook emphasizing its phallic-aggressive aspect. Transformed effects of the incidentally presented cow passage included, for example, an increase in passive and nurturant interpersonal relationships.

The role of the drive relatedness of incidental stimuli is of particular interest. Drive representations are the principal thought contents that undergo symbolization, and drives are principal activators of defensive and disguise processes. In psychoanalytic theory, the relationship of day residues (including incidental stimuli) to unconscious impulses is critical to an understanding of their use in the dream (Freud, 1955). The present study is specifically concerned with variations in the effects of incidental and focal stimuli as a function of the drive relatedness of the stimuli. Both drive and nondrive stimuli were given under incidental and focal conditions and the effects in all cases compared. The issue explored was whether or not the results obtained earlier with symbolic drive stimuli (transformed effect of incidental and direct effect of focal) would be repeated with an explicit (nonsymbolic) drive stimulus and with a neutral stimulus. It was hoped that the findings would shed some light on the relative importance of drive relatedness and of incidentalness in producing transformations of the stimulus content.

METHOD

Subjects (*Ss*) were 24 male undergraduates seen individually in one hour sessions. Each first wrote two stories. Then, two stimuli were presented simultaneously: a "focal" one that *S* read and an "incidental" one, produced in an adjoining room, that was experienced by *S* as irrelevant "noise." The stimuli were verbal descriptions of (*a*) the dependent tie of the male child to his mother (the "drive" stimulus) and (*b*) the workings of a large-scale industrial economy (the "neutral" stimulus). Each was incidental for half the *Ss*. Following this, *Ss* made up two more stories.

An oral, passive, mother-child theme was selected for the drive stimulus because of its universal significance in personality formation. It read as follows:

> Dependent upon his mother from the moment of his birth—this is a man's infancy. Still often feeling helpless in adulthood—this is a man's secret. Growing up is not easy. At the very beginnings of life, all people receive the protecting, cradling, feeding love of their mother. You too long ago felt warm and secure in your mother's arms. You too have sucked ravenously until your belly was full with soothing warm milk. You too have experienced the pleasures of relying completely upon your mother who fulfilled your every wish. And then you find that growing up is not easy. In spite of growing older, the tie to mother remains. Somehow wishing to be mothered even by other women in your adult years. The little boy within, always hiding behind

[1] This study is one of a series supported by a grant from NIMH (Project Code Number: M-782, Cognitive Processes below the Level of Awareness). Principal investigators: George S. Klein and Robert R. Holt.

the appearance of the growing up man. Feelings of helplessness, of aloneness. The wish to be taken care of, and trying to hide this wish. Often feeling like a little boy. Wanting mother.

In the neutral stimulus, an attempt was made to minimize libidinal and aggressive drive implications. Though all such implications could not be fully eliminated, the stimulus used was relatively free of drive content. It read as follows:

Machines have taken a place of enormous importance in the workings of present day civilizations. It is amazing to think of how much machines and their products enter our lives every day. From morning to night, every person makes some use of the products of industry. For example, transportation by car, plane, or boat is taken for granted today. Communication by telephone, printing press, or radio has become a natural part of our lives. The worlds of entertainment, commerce, government, and science all are largely based upon machine products. To a great extent, modern industry and the present day economy require the continued use of machine products by the public. Employment for millions of people can be found in the many factories throughout the country which manufacture such products. The demand for new goods is a constant spur to invention. The economic wealth of a nation is indicated by its accomplishments in industry.

The two stimuli differ in ways other than their drive relatedness since they were written to make the drive stimulus more forceful in general. The oral drive stimulus has more intrinsic interest. It describes a relation of person to person while the other describes a relation of person to machine. There are also differences in writing style. The drive vs. neutral difference is, however, certainly a prominent one and the other differences (interest, relationships) are in the main intrinsically related to it.

The procedure was as follows: *S* and the experimenter (*E*) met in an office where extraneous noise was minimized. *S* had been told only that the study had to do with "imagination and some other things." To begin with, *S* was given a blank card and told to imagine a picture on it and to make up a story about the picture. This was repeated for a second story, with the additional instruction that the story be different from the first one. Instructions were given to *S* in writing to eliminate any need for an inquiry, and no mention was made of including people in the stories (to avoid bias in favor of the drive stimulus). Stories were handwritten by *S* to a length of ½ page, as suggested by *E*.

Following the storywriting, *S* was told that he would get a rest from imagination tasks while *E* got a measure of his "ordinary reading rate" when reading "just for comprehension." Instructions were for *S* to read a passage silently several times, announcing each time a reading was completed so that *E* could note the time. *S* was encouraged to read at a steady rate even as he got increasingly familiar with the material after the first reading; he was told that the aim of the study was to explore "consistency" in reading rate. *S* was given either the drive or the neutral passage to read. Ten seconds after he began reading, the other passage was broadcast from an adjoining office. The reading instructions succeeded in keeping *S*s well focused on the reading task even though the incidental stimulus was quite audible. *E* had *S* read the passage repeatedly until the broadcast of the incidental stimulus was completed; 18 *S*s read the focal stimulus three times, 4 read it four times, and 2 read it five times.[2] After the reading, *E* briefly and casually reviewed the salient points of the passage for *S* (to maximize its focal quality).

The incidental stimulus was tape recorded by a male voice (not *E*'s) and played at a constant volume; it could be heard in the experimental room at close to ordinary speaking level. The incidental stimulus took under 2 minutes to read, and it was read once. The various devices used to increase the likelihood that *S* would experience the incidental stimulus as irrelevant are described in Pine (1960).

Following the reading task, *S* was told that he would now "return to the imagination experiment" and was instructed to write stories once again. Two more stories were obtained, under conditions identical with those described earlier; *S* was asked for a different story each time.

Memory and free association material were next collected (tape recorded). *S* was asked to recall all he could of the material he had read. He was then given a recognition test consisting of 29 pairs of closely matched phrases, with one member of each pair having appeared in the passage; *S* checked off the one in each pair that he thought was correct. Next, *S* was given the focal passage to read through once more and was then asked to free associate to it (for approximately one minute), saying anything that came to mind about the passage or about his own life. Finally, *S* was reminded of the interfering "noise" that he had heard earlier; he was then given the same recall, recognition, and free association procedures for the incidental stimulus.

Generally, *S*s were quite surprised when asked about the incidental stimulus; by and large they had considered it irrelevant, and their memory for it was poor.

RESULTS

Memory

Recall scores were assigned each *S* by dividing each stimulus into information units and scoring the number of units recalled. The recognition score was simply the number of phrases that were correctly identified from the 29 pairs. Average scores and comparisons of recall and recognition for incidental and focal stimuli appear in Table 1.

Recall and recognition of the incidental stimuli are generally low. Recall and recognition of the focal stimulus are significantly higher in each comparison, but still not at a very high absolute level. This is partly because

[2] There was no relationship evident between number of readings and nature of effects.

TABLE 1
MEMORY FOR INCIDENTAL AND FOCAL STIMULI

	Drive		Neutral	
	Incidental	Focal	Incidental	Focal
Recall:				
Total number units	38	38	40	40
Average units recalled[a]	1.2	9.8	0.8	9.3
Range	0–5.5	7.5–13	0–4.5	5–18
t	10.83		7.20	
p[b]	.001		.001	
Recognition:				
Total number units	29	29	29	29
Average units recognized[c]	15.6	23.5	18.4	20.7
Range	9–21	18–26	15–21	15–25
t	5.91		1.77	
p[b]	.001		.05	

[a] Full credit given for near perfect recall of a unit; ½ credit for partial recall.

[b] One-tailed tests.

[c] The average recognition by chance alone would be 14.5 units in the forced-choice situation with 29 pairs of phrases.

Ss were not pressed to give everything they could recall on the free recall, and because, in the recognition tests, the matched pairs of sentences were so similar that confusions were made even though the passages had been read focally.[3]

Stimulus Effects

All stories were studied intensively in an attempt to develop scoring categories appropriate to the raw data. *E* went through half of the stories noting all possible derivatives of either the drive or the neutral stimulus (whether a single word or a broad theme). Indications of direct as well as transformed emergence of stimulus content were recorded. The derivatives were then grouped into the categories that seemed best to describe them (separately for the two stimuli). Finally, *E* rated the other half of the stories by placing possible derivatives directly into the already formed categories.

The rater knew neither which stories were pre and which post nor which stimulus was incidental and which focal. (*Ss* had written out

[3] The memory scores showed no consistent relationship to the stimulus effects to be described below.

TABLE 2
CATEGORY TITLES AND RATINGS FOR EFFECTS OF THE DRIVE STIMULUS

Catagory Titles and Groups	Incidental[a]		Focal[a]	
	Pre	Post	Pre	Post
Direct:				
1. Significant stimulus words used in stories with initial meaning retained (e.g., "he grew *older*")	14	6	16	12
2. Assorted specifics from the stimulus used directly (e.g., eating, birth, social dating, tie to parents, etc.)	42	37	39	71
3. Childhood (and oral) objects and activities (e.g., dollplay)	8	8	6	11
Total	64	51	61	94
Middle:				
1. Phrases parallel to (not identical with) stimulus phrases and used in similar way (e.g., "he enjoyed himself greatly"—from, pleasures of infancy theme)	2	6	2	2
2. Oral aggression (e.g., "lion hunters heard a *roar*")	1	1	1	3
3. General theme: personality (e.g., story of psychology lecture)	1	2	0	2
4. Egocentrism: world centered on self (e.g., daydream—self as hero)	4	2	0	1
5. Miscellaneous: related to various stimulus qualities in various ways	101	92	103	94
Total	109	103	106	96
Indirect:				
1. Significant stimulus words used in stories with altered meaning (e.g., "*full* of life"—from, belly full)	2	3	3	1
2. Phrases parallel to (not identical with) stimulus phrases and used in altered ways (e.g., "he was a hard worker"—from, growing up is not easy theme)	2	2	1	2
3. Oral words used with nonoral meanings (e.g., *rock* and roll)	1	7	2	3
4. Condensations: story words combining two or more stimulus words or concepts (e.g., a "little women" story—from, young and female)	0	1	1	1
5. Puns: story words viewed as plays on stimulus words (e.g., "he felt a *minor* hesitation")	6	7	7	4
6. Slips: errors in story writing in which stimulus content comes through (e.g., "he couldn't *loose* the game"—from, theme of inability to break tie to mother)	4	2	1	1
7. Negations and reversals of stimulus content (e.g., "it was a *cold* and *uncomfortable* day")	15	11	7	10
8. Symbols (of mother, of birth) (e.g., diving into water)	36	24	24	11
9. Concrete representations of oral acts or feelings (nonuniversal symbols) (e.g., "he was *attached* to his seat and couldn't move")	9	3	1	1
Total	75	60	47	35
Grand total	248	214	214	225

[a] Total number of ratings per category for 12 *Ss*.

stories on precoded pages that were shuffled, typed, and re-coded before being seen by any rater.) Thus, at the time stimulus "derivatives" were selected, the rater had no way of knowing whether in fact these were in pre or post stories; his global guesses of which stimulus had been incidental were at a chance level. Sample reliability studies were carried out by two separate raters—each one rating half the stories for a single category. One of the categories selected included relatively direct stimulus effects (reference to social dating—a derivative of the drive stimulus) and the other (oral symbols) included relatively indirect effects. With 24 *S*s, rho between the ratings by *E* and the reliability rater for the dating category was .76 and for the symbol category was .79. Brief category titles and some examples are presented in Tables 2 and 3.

Although grouped in the tables, the "direct" and "indirect" categories are conceived of as extremes of a continuum. Direct categories are defined by the use of stimulus material in the stories in much the same form and content as it appeared in the stimulus; the content is such that a layman could readily recognize the link to the stimulus. Indirect categories were formed and grouped on the basis of clinical impression guided by psychoanalytic theory of the forms of thinking in the primary process (Freud, 1955). They all reflect transformations of the stimulus content based upon fluid associations, word play, and processes of symbolization.[4] The middle categories, a residual group, were those which did not fall as clearly into one or another of the extreme groups. The categories are not mutually exclusive. A single story could get ratings for direct and indirect categories for both stimuli.

In the following, four sets of findings are presented, deriving from: (*a*) incidental drive stimulus, (*b*) incidental neutral stimulus, (*c*) focal drive stimulus, and (*d*) focal neutral stimulus. Since each *S* got two stimuli, the analyses are not based on four independent groups, but on two. Nonparametric tests (two-tailed) were used to study group differences and pre-post differences in the stories. Comparisons were made for the over-all groupings of direct and indirect categories and for the

[4] "Direct" and "indirect" for the drive stimulus were not defined in terms of the directness of expression of oral content, but in terms of the directness of expression of stimulus content—oral or not. Linguistic, as well as drive, links to the stimuli were rated. Later exploratory analyses in terms of the handling of oral dependent drive content per se showed no results.

TABLE 3
CATEGORY TITLES AND RATINGS FOR EFFECTS OF THE NEUTRAL STIMULUS

Category Titles and Groups	Incidental[a]		Focal[a]	
	Pre	Post	Pre	Post
Direct:				
1. Significant stimulus words used in stories with initial meaning retained (e.g., "he went to an *employment* office")	4	4	0	6
2. Assorted specifics from the stimulus used directly (e.g., transportation, work, communication)	83	74	62	72
3. Extensions of assorted specifics from the stimulus—extensions consistent with the initial content (e.g., taxicabs rather than transportation)	32	21	26	24
4. Commentary on the machine age (e.g., *Brave New World* theme)	0	0	0	1
Total	119	99	88	103
Middle:				
1. Phrases parallel to (not identical with) stimulus phrases and used in similar way (e.g., "everybody was working"—from, employment-for-millons theme)	1	6	4	5
2. Personalization: stimulus content used in self-reference context (e.g., "I see myself as a scientist")	0	2	5	
3. Miscellaneous: related to various stimulus qualities in various ways	9	5	24	17
Total	10	13	33	26
Indirect:				
1. Significant stimulus words used in stories with altered meaning	0	0	0	0
2. Phrases parallel to (not identical with) stimulus phrases and used in altered ways (e.g., "the warrior thrust his sword"—from, spur-to-invention)	1	0	2	0
3. Altered use of specific stimulus content suggesting out-of-context associations (e.g., story of baseball team—from, many-people theme)	8	7	11	12
4. Puns: story words viewed as plays on stimulus words (e.g., "time *spent* in school"—from, role of consumer)	1	7	2	5
5. Slips: errors in story writing in which stimulus content comes through (e.g., "the fish *brakes* water"—from, machine and transportation themes)	3	0	0	0
6. Negations and reversals of stimulus content (e.g., a medieval setting—from, present-day theme)	7	2	3	6
7. Concrete representations: symbol-like representations of the stimulus content	0	0	1	0
Total	20	16	19	23
Grand total	149	128	140	152

[a] Total number of ratings per category for 12 *S*s.

total of all categories. Tables 2 and 3 give the total numbers of ratings for all categories and category groupings. Individual scores, from which rank orders were formed, were derived from each *S*'s number of ratings in the categories.

Incidental drive stimulus. In a previous study (Pine, 1960), it was found that an incidental symbolic drive stimulus emerged in indirect ways in subsequent stories. Effects of an incidental explicit (nonsymbolic) drive stimulus were studied here. The total number of ratings on the two pre and the two poststories were computed for each *S* and ranked for the direct, the indirect, and the total drive stimulus categories. Pre to post changes were then compared by means of the signed rank test (Siegel, 1956). Instead of the expected increase in indirect derivatives, there was a significant decrease in the total number of derivatives ($p = .04$). Though the decrease does not appear in every category (see Table 2), it appears in the total for each subgroup of categories (the subtotals alone being in no case statistically significant). Difference scores (changes from pre to post stories) for the drive stimulus derivatives were also computed for each *S* and ranked; *S*s who had received the drive stimulus incidentally were compared to *S*s who had received it focally. A Mann-Whitney test (Siegel, 1956) showed the difference for the total number of derivatives to be significant ($p = .04$), with a decrease in the drive incidental group and an increase in the drive focal group.[5]

Incidental neutral stimulus. To ascertain whether or not the neutral stimulus produced an effect, an analysis exactly parallel to the above was carried out (excepting that categories for the neutral stimulus were used). No significant pre-post differences were obtained on direct, indirect, or total categories for the group of *S*s that received the neutral stimulus incidentally; nor were there any significant differences between groups of *S*s (neutral-incidental vs. neutral-focal). There were fewer derivatives rated in total for the neutral than for the drive stimulus (compare Tables 2 and 3); this reflects the ease with which a clinically oriented rater finds oral-related content in a TAT-like story, and is not a true stimulus effect.

Focal drive stimulus. In the earlier study (Pine, 1960), a focal symbolic drive stimulus seemed to emerge in direct ways in subsequent stories. Present findings with the focal explicit drive stimulus are similar. There was an increase in direct derivatives of the drive stimulus from the pre to the poststories in the group of *S*s that received the drive stimulus focally (signed rank test; $p = .054$). Difference scores (pre to post changes) across groups (drive focal vs. drive incidental) also showed a significant difference (Mann-Whitney test; $p = .04$) with an increase in direct derivatives in the drive focal group and a decrease in the drive incidental group. No other changes from the pre to the poststories for the *S*s in the drive focal group were statistically significant.

The question arose early whether explicit drive stimuli, if threatening to *S*, would undergo some disguise as a function of defense even when focally presented. With this in mind, an oral-passive drive stimulus had been selected as one that would be likely to be threatening for the *S*s used—a group of college freshmen and sophomores presumably in the midst of an adolescent dependence-independence crisis.[6] Though disguise processes (e.g., transformation of derivatives) were not evident, the data do suggest some defensive operations. The free associations to the drive stimulus were rated for anxiety; two raters, neither one aware of the data on stimulus effects, carried out the ratings jointly. The rank order for anxiety correlated inversely, though not significantly, with effects for the *S*s in the drive focal group; rho ($N = 12$) was $-.25$ with direct effects, $-.41$ with indirect effects, and $-.47$ with total effects. The trend is, the more anxiety, the less of any effect. In addition, it was striking that most of the in-

[5] The decrease is not a result of changes in story length, which remained constant for the pre- and poststories (by *E*'s instructions to *S*). Nor is it a result of a shift away from explicit oral content (which did not change) towards content related to the neutral stimulus (which also did not change). The change was predominantly in terms of language usage (see Footnote 4), which decreased in its relatedness to the drive stimulus in the poststories as compared to the baseline of the prestories.

[6] Free associations to the drive stimulus suggest that it was indeed somewhat threatening; the associations varied, not from pleasant to unpleasant, but rather from effective to ineffective handling of essentially unpleasant content.

crease in direct derivatives from pre to post-stories reflected the cast of characters in the drive stimulus (women, boys—not necessarily in a relationship to each other) and the adult relationships suggested in the stimulus (seeking adult ties—marriage, dating); there was almost no increase in (more threatening) direct derivatives concerning the dependent relationship of the infant to its mother.

Focal neutral stimulus. Few if any effects of a focal nondrive stimulus were expected to emerge in the stories (since unthreatening material in full awareness is ordinarily kept from spilling over onto other thought products). Analysis of derivatives showed no significant differences in pre-to-post or in cross-group comparisons.

Since the rating categories used were empirical ones and in no sense exhaustive, there remains a possibility that other categories would have shown different results. This problem is especially true for derivatives of the neutral stimulus; less theory is available to provide guidelines for the formation of "indirect" categories there. Comparison of Tables 2 and 3 indicates that there were in fact many fewer indirect derivatives rated for the neutral than for the drive stimulus, and many more direct derivatives. The data were analyzed in sufficient detail so that one can feel relatively certain that no gross group effects were missed. Nevertheless, one can neither rule out nor confirm the possibility that both stimuli, in either focal or incidental conditions, had a variety of idiosyncratic effects based on stimulus-personality interactions not controlled in the design.

DISCUSSION

In an initial study of the contrasting effects of incidental and focal stimuli (Pine, 1960), results indicated an increase in transformed derivatives of the incidental symbolic drive stimulus in subsequently produced stories. Bach (1960), using a quite dissimilar design, reports the same essential finding. In his study, the words "room" and "water" presented subliminally or incidentally led to mother-related word associations—an effect quite distant from the original stimulus words but in accord with their conceptual value as mother symbols.[7] Given the phenomenon of transformed effects of symbolic drive stimuli, one of the most curious findings of the present study is the *decrease* in *all* derivatives with an explicit drive stimulus.

This result is in line with other studies that point up the significance of stimulus differences (notably symbolic vs. explicit drive) for the phenomenon. Klein, Spence, Holt, and Gourevitch (1958), comparing subliminal stimuli of symbolic and explicit drive content, found group effects of the symbolic stimulus (the response options did not differentiate between direct and transformed effects) but a more variable pattern (individual consistency but no group effects) for the explicit drive stimulus. Eagle (1959), using an explicit drive stimulus presented subliminally, found stimulus effects moving both towards and away from the stimulus content. Thus, it may be that the effects of an explicit drive stimulus are more variable than the (typically transformed) effects of symbolic drive stimuli.

This variation of effects and the symbol-nonsymbol differences mesh in an interesting way with a variety of recent suggestions (Fisher & Paul, 1959; Klein & Holt, 1960; Pine, 1960) that marginal stimuli are governed more by the laws of primary process (nonlogical, drive dominated thought) and focal stimuli by the laws of secondary process (rational, adult thought—cf. Freud, 1955). Symbolization disguises drive content so that its threat is not felt (as in the dream). On the fate of threatening thought content in the psychic apparatus, Freud (1955) writes: "As a result of the unpleasure principle, then, the first ψ-system [primary process] is totally incapable of bringing anything disagreeable into the context of its thoughts. It is unable to do anything but wish" (p. 600). He goes on to say that this makes much material that is necessary for coping with reality unavailable to thought. The secondary process develops as a way of making threatening thought contents available so that adaptive behavior can be maximally effective. Yet, "the second system

[7] The present writer has just repeated this finding in still a different design. The word "cow," presented subliminally to a group of *Ss*, led to an increase in mother-related, oral-passive content in subsequent images—but *not* to an increase in images related to cows, farm animals, food animals, or the like.

[secondary process] can only cathect an idea if it is in a position to inhibit any development of unpleasure that may proceed from it" (p. 601).

We can hypothesize that a "disagreeable" drive stimulus, incidentally presented, is excluded from thought (reduced derivatives in the stories) in accord with the pleasure principle that guides the primary process stream of thought. This contrasts with the fate of incidental *symbolic* drive stimuli which *are* utilized in subsequent thought products (stories), symbolized drive representations typically being available in primary process thought products (e.g., the dream). Since the assumption is that focal stimuli tend to be governed by the laws of the secondary process, it is striking that *S*s who had the drive stimulus focally gave stimulus derivatives that "inhibited the unpleasure" arising from the disturbing thought content. That is, they did not use the more threatening infantile dependent material, but instead used much material related to social dating—for this sample, a kind of content that is congruent with the interests and activities of their peers, its threat, thus, reduced by its very normality. The hypothesis can be tested empirically that threatening stimuli, marginally presented, lead to effects that include inhibition.

An essential problem in studies of effects of marginal stimuli is the role of consciousness in determining the course a stimulus takes in the mental apparatus (cf. Klein & Holt, 1960). A stimulus is "put in" and "comes out" of the person without his full awareness, intake and output being couched in and, thus, hidden by other intake and output processes going on simultaneously. For the drive stimulus at least, the effects obtained repeat a central aspect of results found earlier (Pine, 1960)—that incidental and focal stimuli work differently. The focal stimulus led to an expression of direct derivatives (as found earlier) and the incidental stimulus led to an inhibition of a wide range of derivatives (both direct and transformed). These data suggest that differences in the course stimuli follow in the psychic apparatus stem from differences in their mode of intake.

Recent studies suggest that differences arise from variations on the response side as well. Thus, a linked pair of studies by Goldberg and Fiss (1959) and by Fiss, Goldberg, and Klein (1960), as well as a study by Goldstein and Barthol (1960), show effects of subliminal stimuli on imaginative processes (image and story production) but not on reality-testing processes (discrimination of stimuli). In the present study, memory (a realistic-adaptive kind of thinking) is clearly superior for the focal stimuli; but this difference does not carry over into more imaginative thinking (story creating). It may be that stimuli that are taken in in a way that evades conscious attention and adaptive relevance follow a meandering course in the mental apparatus, and that imaginative response tasks are more sensitive to picking up the end products of this meandering course than are response tasks that emphasize realistic thinking.

It was hoped that the study would illuminate the respective roles of incidentalness and of drive relatedness in determining effects of incidental stimuli. Zuckerman (1960) reports effects of neutral stimuli, subliminally presented. At first glance, present results would seem to contradict Zuckerman's; since an incidental stimulus had an effect when it was drive related but not when neutral, it might be concluded that drive related marginal stimuli stand a better chance of affecting thinking.

Two important methodological considerations modify this conclusion, however, and raise interesting questions in themselves. The first stems from the fact of *simultaneous* presentation of the focal and incidental stimuli. The drive stimulus is not only more drive related but also more *interesting* than the neutral stimulus. The lack of effect of the neutral stimulus presented incidentally may be a function of *S*s' simultaneous attention to a more absorbing focal stimulus; in contrast, the incidental drive stimulus was given in the context of attention to a relatively more boring stimulus.[8] Activities going on in full awareness at the time when incidental stimuli impinge on *S* may, thus, be partial determinants of the effect of the incidental stimuli. To test this hypothesis, one would have to present an incidental neutral stimulus along with an uninteresting focal stimulus.

[8] I am indebted to George S. Klein for pointing out this problem.

The second qualification comes from the role of the congruence (Eagle, 1958) between the response task and the effects to be observed. Congruence of the response task with the stimulus was perhaps greater for the drive than for the neutral stimulus; TAT-like stories are better potential carriers of drive content than of industrial themes (witness the smaller number of ratings for the neutral stimulus). This observation parallels the suggestion raised earlier that the response task must be sensitive to the kind of effects expected.

SUMMARY

The present study continues previous research on the contrasting effects of incidental and focal stimuli upon subsequent thinking. The significance of stimulus content (notably drive-nondrive) for these contrasting effects was specifically explored. The larger theoretical context grows from concern with the role of consciousness and attention in determining the course a stimulus takes in the mental apparatus.

Twenty-four male *S*s were seen individually for a study of "imagination." Each told four stories to a blank card—two pre-experimental (control) and two post-experimental. The intervening experimental condition was the simultaneous presentation of two stimuli. One, the "focal" stimulus, was a written passage that *S* was asked to read as a measure of his "ordinary reading speed." The other, the "incidental" stimulus, was a similar passage broadcast through the wall from an adjoining room and characteristically experienced by *S* as irrelevant "noise." One passage was drive related (the dependent tie of infant to mother) and the other was more neutral (a description of an industrialized economy). Each was incidental for half the *S*s (while the other was simultaneously focal).

Main effects were determined by comparisons of the pre and the poststories. Effects were categorized in terms of their direct or transformed relation to the initial stimulus. The incidentally presented drive stimulus produced an inhibition of direct and transformed content that was related to the stimulus, while the same stimulus focally presented led to an increase in specific kinds of direct stimulus content. The neutral stimulus produced no observable effects in either the incidental or the focal conditions.

The results were discussed in the light of current studies of the often unique effects of marginal stimuli (incidental and subliminal) upon contrasting kinds of thought (reality-attuned vs. imaginative). Suggestions regarding the contrasting nature of the effects of incidental and focal stimuli, and the significance of the drive relatedness of incidental stimuli, were drawn from the psychoanalytic theory of primary and secondary process thinking.

REFERENCES

BACH, S. Symbolic associations to stimulus words in subliminal, supraliminal, and incidental presentation. Unpublished doctoral dissertation, New York University, 1960.

EAGLE, M. Influences of subliminal aggressive stimuli upon conscious cognition. Unpublished doctoral dissertation, New York University, 1958.

EAGLE, M. The effects of subliminal stimuli of aggressive content upon conscious cognition. *J. Pers.*, 1959, **27**, 578–600.

FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. Psychoanal. Ass.*, 1957, **5**, 5–60.

FISHER, C., & PAUL, I. H. The effects of subliminal visual stimulation on images and dreams: A validation study. *J. Amer. Psychoanal. Ass.*, 1959, **7**, 35–83.

FISS, H., GOLDBERG, F. H., & KLEIN, G. S. Effects of an undiscriminated stimulus on imagery. New York: New York University, Research Center for Mental Health, 1960. (Mimeo)

FREUD, S. *The interpretation of dreams.* (Originally published 1900) New York: Basic Books, 1955.

GOLDBERG, F. H., & FISS, H. Partial cues and the phenomenon of "discrimination without awareness." *Percept. mot. Skills*, 1959, **9**, 243–251.

GOLDSTEIN, M. J., & BARTHOL, R. P. Fantasy responses to subliminal stimuli. *J. abnorm. soc. Psychol.*, 1960, **60**, 22–26.

KLEIN, G. S. Consciousness in psychoanalytic theory: Some implications for current research in perception. *J. Amer. Psychoanal. Ass.*, 1959, **7**, 5–34.

KLEIN, G. S., & HOLT, R. R. Problems and issues in current studies of subliminal activation. In J. G. Peatman & E. Hartley (Eds.), *Festschrift for Gardner Murphy*. New York: Harper, 1960. Pp. 75–93.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious sought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

KRIS, E. *Psychoanalytic explorations in art.* New York: International Univer. Press, 1952.

PAUL, I. H., & FISHER, C. Subliminal visual stimula-

tion: A study of its influence on subsequent images and dreams. *J. nerv. ment. Dis.*, 1959, **129**, 315–340.

PINE, F. Incidental stimulation: A study of preconscious transformations. *J. abnorm. soc. Psychol.*, 1960, **60**, 68–75.

SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SMITH, G., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

ZUCKERMAN, M. The effects of subliminal and supraliminal suggestion on verbal productivity. *J. abnorm. soc. Psychol.*, 1960, **60**, 404–411.

(Received April 4, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 491-496

THE RELATIONSHIP BETWEEN WORD FREQUENCY AND PERCEPTIBILITY WITH A FORCED-CHOICE TECHNIQUE

JANET A. TAYLOR, DORIS C. ROSENFELDT, AND RUDOLPH W. SCHULZ[1]
Northwestern University

EMPIRICAL relationships between recognition threshold and such properties of the stimuli as prior familiarity or affectivity have been well-established, as well as the interaction of these stimulus variables with motivational variables or subjects' (*S*s') personality characteristics (Jenkins, 1957; Postman, 1953). The usual interpretation of these data has been in terms of perception. That is, differences in threshold between groups of *S*s or types of stimulus material are attributed to the relative ease with which the stimuli are "seen," or in the case of auditory presentation, "heard." These results have been used in an attempt to extend our knowledge of perceptual processes and, by many investigators, to support explanations of personality phenomena based on perceptual theory.

Recently Goldiamond (1958) has challenged the perceptual interpretation of many of these studies on both methodological and empirical grounds. He points out that the performance measure obtained from the recognition threshold technique in which *S* is required (by the method of ascending limits or constant stimuli) to identify a stimulus under perceptually inadequate conditions is particularly susceptible to the influence of variables extraneous to the properties of the presented stimulus. Particularly important are what Goldiamond terms "response biases"; i.e., under conditions in which a stimulus is only partially perceived, *S*s are more likely to guess or say certain responses than others due to past learning, motivational factors, the experimenter's (*E*'s) instructions, etc.

A restatement of the Goldiamond (1958) position, as it is understood by the present writers, may be made via the following examples. If *S* were presented, at a suprathreshold level, with a stimulus pattern such as C-T and instructed to report what he *saw*, he would report C-T; the verbal responses elicited by these particular instructions, it could be agreed, indicate or reflect *S*'s perceptions. The *S* could also be asked to fill in the missing letter by saying the first complete word that occurred to him (e.g., CAT). One would *not* conclude in this instance that *S*'s behavior reflected his perception (e.g., that he saw CAT rather than C-T). It is quite reasonable to assume that in this type of situation, such variables as the prior frequency of usage of the various response alternatives, their affective connotations, *S*s' motivation and set, etc. influence which response alternative is elicited. Were these empirical relationships to be obtained, one would *not* conclude that they constitute evidence that these variables determine perception. Rather they would be interpreted in other terms.

The technique of obtaining recognition threshold by the ascending method of limits contains elements of both of the procedures described above, one of which elicited behavior that can be agreed to reflect perception and the other of which does not. In the typical experiment *S* is asked to report after each exposure not merely (or sometimes, not even) what he *sees* but also to *guess* at what the word is on the basis of whatever cues are available to him. Obviously, *S*'s verbal responses at subthreshold levels of exposure are determined to some degree by his perceptions, i.e., they are determined by the immediate properties of the stimulus. But it is also well-known that recognition of complex stimulus patterns like words is not all-or-none (Bricker & Chapanis, 1953); partial perceptions of increasing degrees of completeness and accuracy occur as exposure values increase toward suprathreshold levels. When *S* guesses at a whole word on the basis of a partial perception (thus paralleling the suprathreshold study described above) it is difficult to maintain that *S*'s verbal behavior indicates only what he sees or perceives.

With respect to frequency of prior usage

[1] Now at the State University of Iowa.

several studies indicate that this variable does influence *S*s' verbal responses when partial perceptions are known and held constant. In two of these (Goldiamond, 1959 unpublished; Goldiamond & Hawkins, 1958) *S*s were shown blank slides that allegedly contained stimulus items. "Thresholds" were established by counting the number of exposures necessary for *S* to guess the response *E* had previously decided upon as being "correct." These pseudothresholds were found to be related to prior familiarity with the stimulus materials in the same manner as in threshold studies in which stimuli were actually presented (e.g., Howes & Solomon, 1951; Solomon & Postman, 1952). A similar phenomenon has also been reported by Zeitlin (1953) who manipulated the reinforcement history of the stimuli. The results of such studies strongly suggest that the empirical relationship between recognition threshold and at least one variable, frequency of prior usage, is due at least in part to guessing preferences. At lower exposure values that provide only minimal information (and thus permit a number of response alternatives) *S* will tend to say (guess) familiar words more often than less familiar ones. This phenomenon has been reported by McGinnies, Comer, and Lacey (1952) and by Rosenzweig and Postman (1957). Should the actual stimulus item happen to correspond to his guess, *S*'s "threshold" is low. Should the stimulus happen to be a less familiar word, additional exposures, which provide additional information that narrows the range of possible responses, are required before the correct response is elicited.

In the present investigation an attempt was made to show a relationship between frequency of prior usage and verbal report, using a procedure better able than the recognition threshold method to minimize the biasing effects of guessing preferences. The experimental arrangement was a modification of the forced-choice psychophysical technique employed by Blackwell (1952) and Goldiamond (1954).[2] As employed in the present study, *S* was presented with several words simultaneously and required to indicate the spatial position of one of the words which had previously been identified by *E*. This method restricts severely the verbal responses that are elicited; bias due to any differential preference of one response (spatial position) over others can be minimized by assigning high and low frequency words equally to all spatial positions.

[2] The forced-choice technique has been used by several investigators (Blum, 1954; Dulaney, 1957; Nelson, 1955) in studies of the affective nature of the stimulus materials. In whole or in part, however, these investigators required *S* to indicate the location of the "most outstanding" stimulus, a procedure which Goldiamond (1958) contends is subject to the same criticism as the ascending method of limits.

Experiment I

Method

Subjects. Undergraduate psychology students at Northwestern University were tested in groups with the forced-choice method. The study was replicated three times, the groups containing 17, 18, and 19 *S*s, respectively. Four additional *S*s were tested but their records were excluded because they failed to reach a performance criterion (a minimum of 75% correct responses at the highest illumination level).

Apparatus and materials. Eight pairs of six-letter nouns (a total of 16 individual words) were selected, four pairs with high frequency of usage and four with low, as indicated, respectively, by Thorndike-Lorge (1944) counts of 100 or more and 7 or less per million. Within each pair, each word began with a consonant and no common letter appeared in the same position. An attempt was also made to equate the frequency with which individual letters appeared in the high and low frequency lists.

Slides were prepared for the word pairs in which one word appeared directly above the other. Two such slides were made for each pair, the spatial position of the words being reversed. The slides were projected onto a screen at a constant .10 seconds duration interval and at six different levels of illumination. Illumination was varied by having the power supply (110 volts 60 cycle ac) for the Keystone slide projector, which contained a 1000 watt lamp, pass through an Ohmite rheostat.

Procedure. *S*s were tested as a group in a semidarkened room, being seated 10 to 18 feet from the screen, in rows of five. Ninety-six stimulus presentations were given by the method of constant stimuli. Prior to each exposure one of the pair of words was announced by *E* and *S*s were required to indicate on printed data sheets the spatial position, top or bottom, in which the given word appeared. The pairs were presented at 6 illumination levels, *S*s being asked to locate each of the 16 words once at each level. Each individual word to be identified appeared equally often in the two spatial positions, as did the 16 different words within each illumination level. The order of the 96 exposures was random with respect to order of word, word frequency, illumination level, and correct spatial position.

Results and Discussion

For each group separately, the mean percentage of correct responses for high and low frequency words at each illumination level was

first obtained. Inspection of these data revealed that a full psychophysical function was obtained in each group; further the curves for the high and low frequency were essentially parallel. For this reason, the data chosen for statistical comparison were the means of percentage of total correct responses for the two types of words. As may be seen in Table 1, the means of the high and low frequency lists are essentially equal for one of the groups but greater for the high frequency words in the other two. Matched-pairs t's computed for each of the groups failed to reach the .05 level of significance although that for Group 3 fell just below it. A similar t for all groups combined yielded a value of 1.81 ($.05 < p < .10$). Thus there was no clear-cut evidence that the positions of high frequency words were more often correctly identified than those for low frequency words.

Although the two lists of words employed in the present study differed widely in Thorndike-Lorge (1944) word count, it could be argued that all of the words were sufficiently familiar to college students to prevent the appearance of differences in perceptibility. To check this possibility recognition thresholds were determined tachistoscopically for the 16 individual words by the usual ascending method of limits. The results obtained from 15 *Ss* revealed highly significant differences between the high and low frequency lists ($t = 5.15$, $p < .001$). Thus the failure to find significant differences with the forced-choice technique could not be attributed to the relative frequency levels of the stimuli.

A second factor that should be considered in evaluating these results is the arrangement of the word pairs with respect to common letters. Since some decision had to be made as to the number of common letters in the same position in each word that would be permitted, the least arbitrary number, zero, was chosen. In effect, then, *S* could be correct in his spatial identification if he perceived correctly the position of one letter in either of the two words. Although *Ss* were not informed of this arrangement, many undoubtedly based a number of their responses on extremely limited information. Since the frequency of individual letters was held constant between the two lists of words, it is perhaps not surprising that only minimal differences were found between high and low frequency words.

Were the number of stimuli increased, the probability of the occurrence of correct responses based on such limited information would decrease sharply. An additional experiment was therefore performed, one in which *Ss* were presented with sets of four high or low frequency words.

TABLE 1

MEANS OF PERCENTAGE TOTAL CORRECT RESPONSES AND t VALUES FOR THE HIGH AND LOW FREQUENCY LISTS

(Experiment I)

Frequency	Group 1 ($N = 17$)		Group 2 ($N = 19$)		Group 3 ($N = 18$)	
	M	σ	*M*	σ	*M*	σ
High	69.36	9.90	77.85	11.39	74.31	13.10
Low	69.73	9.65	76.54	23.12	70.25	14.60
t	.16		1.03		2.01	

EXPERIMENT II

Method

Subjects. *Ss* were 66 students enrolled in introductory psychology classes at Northwestern University. The *Ss* were randomly divided into two equal groups, each being given a different but parallel set of stimulus materials. Thus, in effect, the experiment was replicated twice. All *Ss* were tested individually.

Materials and apparatus. Two lists of 16 five-letter nouns were prepared, both according to the same rules. Each list contained eight high frequency and eight low frequency words, as defined by Thorndike-Lorge counts of 100 or more and 4 or less per million, respectively. Each list was further subdivided into two sets of four high or four low frequency words. Within each set, each word began with a consonant and no common letter appeared in the same position. The frequency with which individual letters appeared in the high and low frequency lists was equated as closely as possible.

Slides of each of the word sets were prepared, the words appearing in capital letters within a $1'' \times \frac{3}{8}''$ rectangle. Four different slides were made for each set by selecting one order of the four words (UR, LR, UL, LL) and rotating the position of the words for each of the other three orders so that each word appeared once in each spatial location.

The slides were presented by means of a Gerbrands Mirror Tachistoscope. Illumination was held constant and the materials exposed at eight duration levels: 10, 20, 30, 40, 50, 60, 80, and 100 milliseconds. Both exposure and pre-exposure fields were white and of equal illumination. A fixation point on the pre-exposure field marked the center of the rectangle in which the block of words occurred.

Procedure. Each *S* was tested with only one of the

TABLE 2
MEANS OF PERCENTAGE TOTAL CORRECT RESPONSES AND t VALUES FOR THE HIGH AND LOW FREQUENCY LISTS
(Experiment II)

Frequency	Group 1		Group 2	
	M	σ	M	σ
High	72.58	15.53	75.90	13.91
Low	66.09	13.73	72.16	13.03
t	4.61		3.57	

two parallel sets of words, *Ss* being assigned alternately to each set as they appeared to be tested. Each *S* was given 128 stimulus presentations by the method of constant stimuli. Prior to the exposure of each slide, *E* announced one of the four words and required *S* to indicate the spatial position, UR, UL, LR, LL, in which the given word appeared. *S* was asked to identify each of the 16 individual words once at each of the eight duration levels. Each individual word to be identified occurred equally often at each spatial position as did the 16 words within each duration level. The order of the 128 presentations was random, a different order being used for each *S*.

Results

An analysis of mean percentage of correct responses at each duration level indicated that the high frequency words were quite consistently located more accurately than the low frequency words in both groups. This superiority is reflected in the means of percentage of total correct responses shown in Table 2. The t testing the significance of the difference between these means was significant in each group ($p < .01$).

DISCUSSION

In conducting research on the recognition thresholds of verbal materials it is possible to have a purely empirical interest in isolating stimulus and subject variables that contribute to verbal behavior (or more exactly, influence the exposure value at which *Ss*' verbal report is congruent with the presented stimulus). More frequently, however, investigators have used the data from such studies to support hypotheses about "perception," an underlying state of the organism. Perception is thus conceived, implicitly or explicitly, as a hypothetical concept inferred from behavior rather than as a convenient term for labeling the behavior observed in a particular experimental setting.

Investigators who use the term "perception" as an explanatory concept have been handicapped by lack of a generally accepted definition of the concept. Typically (visual) perception appears to refer in some manner to "what *S* sees," although how this is to be inferred from behavior is left unspecified. Perhaps as a result, many investigators have appeared to assume, for all practical purposes, that the systematic variance in verbal reports can be attributed to variations in *Ss*' "perception"; thus all variables shown to influence *Ss*' behavior in the recognition threshold situation are almost automatically interpreted as support for a hypothesis concerning perception. The assumption of a 1-to-1 correspondence between behavior and perception has led to a series of criticisms, each suggesting or demonstrating that a given set of empirical results could be attributed to nonperceptual factors. For example, the tabu word studies were questioned because of the possibility of "response suppression"; *Ss* might see (perceive) tabu words as easily or quickly as neutral words but hesitate to say them, thus producing higher thresholds (Howe & Solomon, 1950). The "see" vs. "say" issue is raised once more in Goldiamond's recent articles, Goldiamond pointing to the particular susceptibility of the response used in the method of limits technique to the influence of all sorts of variables that affect performance via nonperceptual processes.

The forced-choice technique would appear to overcome Goldiamond's methodological criticisms of the method of limits. In the present investigation evidence was obtained that indicated that word frequency is related to verbal report even when response bias is controlled. The survival of the empirical relationship between performance and prior frequency of usage with still another source of extraneous variance reduced or eliminated may give added confidence to the perceptual interpretation. But even in the forced-choice situation *S* is still perceiving partial cues and guessing as to the spatial location of the designated stimulus. Thus the question remains unanswered as to whether perception is influenced by frequency of prior usage or whether fewer partial cues are needed to identify more familiar materials.

Whatever the psychophysical method,

partial perception would seem to be inevitable if complex patterns are to be employed. The search for a "pure" measure of perception would thus appear to be a futile one. The latter conclusion is hardly novel. For example, Postman (1953) has suggested that an inference about a perceptual response disposition can safely be made only if it is based on two or more classes of responses. A similar approach has been taken by Garner, Hake, and Eriksen (1956) who urge a program of convergent operations.

What is sought by multiple measurements is a consistent empirical relationship between performance and the variable under consideration, one that is common to all sets of operations rather than unique. If the selection of the sets of operations is such that only the perceptual hypothesis is adequate to explain the observed consistency of results, the perceptual hypothesis is thus confirmed and all alternate hypotheses (e.g., response suppression, linguistic preferences) are disconfirmed. However, as was indicated above, *all* experimental arrangements which involve testing of hypotheses about the perceptibility of complex stimuli (e.g., words) seem inevitably to evoke overt behaviors that are responses to *partial* perceptions. Thus one does not, by using multiple measurements, rule out all alternatives but the perceptual hypothesis. For example, the results of the present investigation suggest that the empirical relationship between recognition thresholds and word frequency found with the ascending method of limits may not be attributed solely to biases in emitting more familiar responses when a number of response alternatives are possible. But they are still susceptible to a nonperceptual interpretation: fewer partial cues are necessary to identify familiar stimuli. It has been the former source of contamination, it is interesting to note, which has occupied the attention of Goldiamond, Postman, Garner et al. and not the latter.

There is no reason, however, why attempts at verification of perceptual hypotheses concerning threshold data need to be restricted to investigations within this limited area. For example, there is already a body of literature (e.g., Osgood, 1957) concerning the identification of stimulus patterns presented in partial but suprathreshold form. If a variable that has been shown to influence threshold behavior can be shown *not* to influence response identification when partial cues are presented at a suprathreshold level, a firmer basis would be provided to support a perceptual interpretation of the former. Conversely, the variables shown to affect thresholds might be found to operate in the same manner to determine response identification of partially presented stimulus patterns. If such were the case, one might conclude that the unique contribution of the scores of threshold studies that have appeared over the past 15 years has been to demonstrate that short durations of exposure, low illumination levels, blurred carbon copies, etc. are all methods for obtaining partial perceptions.

SUMMARY

A number of studies have suggested that word thresholds obtained from the ascending method of limits are particularly susceptible to linguistic biases and hence may indicate little about perception. In the present investigation a forced-choice technique was employed in which *S* was presented with several words and required to identify the spatial location of a given word. With this technique, which reduces or eliminates response bias, the same positive relationship between frequency of prior usage and performance as had been found with the ascending method of limits was obtained. The implications of a perceptual interpretation of these findings were discussed.

REFERENCES

BLACKWELL, H. R. The influence of data collecting procedures upon psychophysical measurement of two sensory functions. *J. exp. Psychol.*, 1952, **44**, 306–315.

BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **49**, 94–98.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly-perceived stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

DULANEY, D. E. Avoidance learning of perceptual defense and vigilance. *J. abnorm. soc. Psychol.*, 1957, **55**, 333–338.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

GOLDIAMOND, I. The relationship of subliminal perception to forced choice and psychophysical judgments, simultaneously obtained. *Amer. Psychologist*, 1954, **9**, 378–379. (Abstract)

GOLDIAMOND, I. Indicators of Perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55,** 373–411.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuche: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56,** 457–463.

HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57,** 229–234.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41,** 401–410.

JENKINS, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54,** 100–127.

MCGINNIES, E. M., COMER, P. B., & LACEY, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44,** 65–69.

NELSON, S. E. Psychosexual conflicts and defenses in visual perception. *J. abnorm. soc. Psychol.*, 1955, **51,** 427–433.

OSGOOD, C. E. Motivational dynamics in language behavior. In, *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957. Pp. 348–424.

POSTMAN, L. The experimental analysis of motivational factors in perception. In, *Current theory and research in motivation: A symposium.* Lincoln: Univer. Nebraska Press, 1953. Pp. 59–108.

ROSENZWEIG, M. R., & POSTMAN, L. Intelligibility as a function of frequency of usage. *J. exp. Psychol.*, 1957, **54,** 412–422.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *J. exp. Psychol.*, 1952, **43,** 195–202.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, 1944.

ZEITLIN, L. R. A response oriented analysis of the concepts of autism and perceptual sensitization. Unpublished doctoral dissertation, Northwestern University, 1954.

(Received April 4, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 497-503

PRECONSCIOUS PERCEPTION: A RE-EXAMINATION OF THE POETZL PHENOMENON[1]

HAROLD JOHNSON AND CHARLES W. ERIKSEN
University of Illinois

In 1917 Poetzl reported that when his patients were asked to dream about a briefly exposed picture many items not mentioned in deliberate attempts at recall appeared in distorted form in the patients' dreams. Over the years there were several reported confirmations of Poetzl's findings (Malamud, 1934; Malamud & Linder, 1931), but the current revival of interest in this phenomenon has been mainly due to the work of Fisher (1954, 1956, 1957). While these studies have been interpreted as confirming Poetzl's original finding, they have lacked much in terms of methodological sophistication and precision in assessment of the amount of "unconscious recall."

Recently Shevrin and Luborsky (1958) have attempted to establish the "Poetzl phenomenon" on a firmer basis, primarily through improved quantification in the assessment of the unconscious or preconscious recall effects. Subjects (*Ss*) were shown a scene by means of a tachistoscopic exposure and were asked to describe the scene and recall the items contained in it. After a second attempt at deliberate recall they were instructed to dream that night and report the following day for a second session. Since a large proportion of the *Ss* did not bring dreams to the second session, they were asked to conjure up a daydream or image. The *Ss*' images and dreams as well as their intentional recalls were then rated for the presence of items contained in the stimulus picture. These ratings were in terms of conceptual similarity to items contained in the original picture rather than in terms of a criterion of specific identity.

When the dreams and images were corrected for the number of items mentioned in intentional recall, Shevrin and Luborsky (1958) found a significant number of conceptually similar items in the dreams and images over and above that contained in intentional recall. This they ascribed to the effect of preconscious perception of the stimulus items during the tachistoscopic exposure of the picture and the subsequent use of these preconscious perceptions by the primary process in dreams and images.

While Shevrin and Luborsky (1958) introduced some important refinements in quantification of the Poetzl effect, their study lacks several basic controls that are imperative if one wishes to explain the results in terms of preconscious perception and preconscious recall. Since a criterion of conceptual similarity was used in comparing the stimulus picture with the images (a truck in the picture is considered recalled if the dream contains a bicycle, since they are both modes of conveyance) rather than specific identity, a very careful estimation of the chance level at which various kinds of objects appear in dreams and images of *Ss* who have not been exposed to the picture needs to be obtained. The images from these *Ss* would have provided an estimate of the base rate necessary to evaluate whether significant preconscious recall occurred among the experimental *Ss*.

A second source of artifact in their data arises from their use of a cohesive scene as the stimulus picture. Since the stimulus used is a natural representation of the environment, objects contained therein have a high probability of being found together in our everyday experience. This could have resulted in a spurious preconscious recall effect in the following manner. Let us assume that after a tachistoscopic exposure of a scene the only thing an *S* can deliberately recall is the impression of a lake. Now, if shortly thereafter he is asked to form an image, it is not unreasonable to suspect that the image or daydream is in part determined by his associations with a lake. And his associations to a lake, furthermore, would not be expected to be random. It would seem that such things as boats, wharves, trees, birds, etc. would occur with a high probability relative to such items as marbles and calculators. Since

[1]This research was supported by Research Grant M-1206 from the National Institute of Mental Health, United States Public Health Service.

it is also probable that a stimulus picture of a cohesive naturally occurring scene containing a lake will also have some of these objects in it, *S* is likely to receive credit for preconscious recall of objects the presence of which is the result of prior learned associations and independent of any preconscious perception or memory.

There are several methods by which this artifact due to prior learned associations can be eliminated or evaluated in preconscious recall. One method would be the use of a control group that is not exposed to the stimulus pictures. Instead *S*s in this control group are randomly paired with experimental *S*s and each control *S* reads over and studies the original recall of his matched experimental *S*. Images are then obtained from the control group and the images are evaluated for the occurrence of items conceptually similar to those in the stimulus picture that were not mentioned in the deliberate recall of the experimental *S*. A second method would be to use stimulus pictures composed of unrelated objects. One could use a random method of selecting a list of objects and then randomly assigning a given number of them to a stimulus card. This would serve to preclude spurious recall effects due to prior learned associations among the objects.

Finally, the Shevrin and Luborsky study, as have most of the other studies on the Poetzl phenomenon, made use of *S*s who had an unspecifiable amount of prior contact and association with the *E*s. This uncontrolled association makes it impossible to specify the exact nature of the experimental conditions and instructions to the *S*. One cannot know the degree to which *S*s may be contaminated by the expectations of the *E*s and the degree of nonexplicit verbal and nonverbal communication that constitutes part of the real instructions given to the *S*.

The present study is an attempt to assess the amount of preconscious recall when these sources of artifact have been eliminated. Using a method similar to that of Shevrin and Luborsky (1958) we have assured not only independence between the judges' ratings of the unintentional recall material and their knowledge of the experimental conditions, but in addition have used a control group to assess the base rate with which conceptually similar objects appear without the pictorial stimulation. A control group was also planned to assess the effects of prior learned associations among the objects in the pictures but this proved to be unnecessary. The unconscious recall effect was eliminated completely when correction was made for base rate.

METHOD

Procedure

Three groups of *S*s, two experimental and one control, were employed. The *S*s were assigned to the three treatment groups alternately in terms of the order of their appearance for the first experimental session. All experimental *S*s were seen individually at both sessions and in the first session approximately 5 minutes was spent in conversing with *S* in order to establish rapport. The nature of a tachistoscope was explained to them and they were told that on the count of three a picture would appear briefly in the tachistoscope. The stimulus was then exposed for .02 seconds at an illumination level of 1.80 foot-candles as determined by a McBeth Illuminometer. For *S*s in Experimental Group I the stimulus was an archeological scene (Stimulus A) and for Experimental Group II an early historical waterfront scene (Stimulus B).

Following the exposure *S* was asked to draw and describe as fully as he could what he had seen. Encouragement was given to overcome hesitancy in drawing and to reassure him that drawing talent was not being assessed. Upon completion of the drawing and verbal description the drawing was removed and *S* was then asked to draw and describe again what he had seen. He was asked to start from the beginning and to draw and describe it as fully as possible. Following this, *S* was asked how frequently he had dreams. He was then asked if he would remember and write down upon awakening any dreams that he might have that night, and to bring these dreams with him to the session that was then scheduled for the next day.

At the second session *S* was asked whether he had dreamed and, if so, he was instructed to draw and describe the dream fully. Following the description of the dream *S* was asked to sit back, relax, and close his eyes, and to let his mind wander or daydream until some sort of picture came to mind. He was then asked to draw and describe as fully as he could the mental image. *S*s who did not report dreams were asked to go through the image procedure twice. Following these reports of dream images, the *S*s were asked to draw and describe as fully as possible the picture that had been shown to them the previous day.

The control *S*s participated only in this second session and were treated the same as the experimental *S*s with the exception that approximately 5 minutes of this second session were devoted to the establishing of rapport, and, of course, the third recall attempt was not required.

Apparatus and Stimuli

The stimulus pictures were presented in a Dodge mirror tachistoscope. The size of the pictures was 5″ × 8″ at a distance of 24″ from the eye. The archeological scene (Stimulus A) used with Experimental Group I depicted a large pyramidal structure in a semi-tropical setting. The scene contained a great deal of vegetation suggesting a dense surrounding jungle. Human figures appeared as only a minor part of the picture. The second picture (Stimulus B) was a waterfront scene depicting a Phoenician city in which a number of dark skinned persons are prominently engaged in construction work while buildings, water, and boats appear less prominently in the background. The experimental sessions were recorded on tape and the recall, dream, and image descriptions were transcribed for evaluation.

Subjects

Forty-five male undergraduates randomly drawn from the introductory psychology course at the University of Illinois were evenly divided among the three treatment groups. None had previously participated in experiments of this nature and they had had no previous contact with the *E*.

RESULTS

Only 10 of the 45 *S*s brought dreams to the experimental session. In order to have a measure that was comparable across all *S*s it was decided to analyze only the first image for each *S*. This decision was abetted by the Shevrin and Luborsky (1958) finding that the preconscious recall effects were more prominent in the image material than in the dreams.

For use in scoring and evaluating the images as well as the intentional recalls, a check list was constructed for each of the stimulus pictures enumerating the items contained in the picture. Stimulus A was broken down into 21 fairly discrete categories and Stimulus B was similarly divided into 29. The drawings and verbal descriptions for both the images and the intentional recalls were evaluated by four advanced graduate students of clinical psychology. Each pair of these judges evaluated the material of two-thirds of the *S*s. Prior to evaluating the protocols, the judges were given a group training session during which scoring criteria were discussed and several practice images and drawings were evaluated. In evaluating the protocols, however, the judges worked independently. Their first task was to rate all of the images on the check list for Stimulus A and then B. After the images had been rated the judge evaluated the descriptions of the stimulus or the intentional recalls. These evaluations were made on the same check lists of items.

In evaluating the images the judges were instructed to rate in terms of conceptual similarity as was done by Shevrin and Luborsky (1958). Thus, a spear that appeared in Stimulus B was credited as being present in the image if an *S* reported any type of weapon in the image. Particular attention was given to this criterion of conceptual similarity during the training session. Although the judges had available to them the two stimulus pictures, their evaluations were made independent of the knowledge of the treatment group to which an *S* had been assigned.

The reliability of the judges' evaluation of the images was computed by means of Pearson product-moment correlations between all possible pairs of the four judges. The average correlation was .66.

In order to facilitate comparison of our results with those of Shevrin and Luborsky (1958) we have adopted their nomenclature for the different recall measures computed. *Intentional recall* consists of all the separate items mentioned by an *S* in his three deliberate attempts to recall the stimulus picture. *Related imagery* items are all the items mentioned in an *S*'s image that are conceptually similar to items in the check lists for either Stimulus A or B. The related imagery score is further subdivided into two scores. First, there is the *preconscious recall* score which comprises the imagined items that are related conceptually to items in the stimulus picture and have not been reported in any of the three attempts at *intentional recall*. Second, the *recalled again* score represents those imagined items that also have appeared in one or more of the intentional recalls. Finally, *unintentional recall* means all the items contained in the *S*'s image.

Since the agreement between judges was less than perfect, two parallel analyses or sets of recall scores were computed. The first set is for the perfect agreement condition in which both judges of a pair agree upon the presence of the item. The second, the incomplete agreement condition, is based upon indication by one or the other judge that the item was pesent.

Our first step in analyzing the data was to

determine whether or not the imagery obtained from the three groups was comparable irrespective of its relation to stimulus pictures. A simple analysis of variance was run on the unintentional recall scores for the *S*s in the three groups. The between groups F of .35 ($p > .10$, df 2 and 42) did not appraoch statistical significance. The comparability of the three groups can be seen in Table 1 where the mean unintentional recall score by group is given. This table also shows the mean intentional recall scores for the two experimental groups.

The comparability of the three groups on unintentional recall is fortunate since it facilitates the comparison of the three groups on the number of imagined items that are conceptuaally similar to the items in the stimulus pictures. For this analysis we computed the related imagery score for each *S* in each of the three groups for Stimulus A and also for Stimulus B. These scores were evaluated by a modified three-way classification analysis of variance (groups, stimuli, and individuals). A separate analysis was performed for the perfect agreement and the incomplete agreement scores. In keeping with the concepts of preconscious perception and preconscious recall it was anticipated that *S*s in Group I would have a greater amount of related imagery to Stimulus A than either Group II or the control group. Similarly, Group II should have a significantly greater related imagery score for Stimulus B than either of the other two groups. This effect should be revealed in a significant group-by-stimulus interaction in the analyses of variance.

These expectations, however, were not confirmed in either analysis. For the items showing perfect agreement among judges the interaction F was .49 ($p > .10$, df 2 and 42) while for the incomplete agreement scoring the F was 1.76 ($p > .10$, df 2 and 42). The only significant effect in either analysis was the between stimulus picture effect for the perfect agreement data ($F = 9.83$, $p < .01$, df 2 and 42). As can be seen in Table 2 this significant effect arises from the fact that all three groups have a smaller number of items in their images related to the items contained in Stimulus B.

TABLE 1
MEAN NUMBER OF ITEMS IN UNINTENTIONAL AND INTENTIONAL RECALLS

Group	Unintentional	Intentional	
Experimental I	11.9	8.4	*5.7*
Experimental II	10.7	9.6	*5.9*
Control	11.9		

Note.—Values in italics represent perfect agreement condition.

TABLE 2
MEAN RELATED IMAGERY IN UNINTENTIONAL RECALL

Group	Stimulus A	Stimulus B
Experimental I	10.4	8.2
	5.0	*3.5*
Experimental II	8.9	8.4
	4.3	*3.6*
Control	10.8	10.9
	5.1	*4.1*

Note.—Values in italics represent perfect agreement condition.

The data in Table 2 in conjunction with the above analyses provide no evidence that the nature of the tachistoscopically presented picture, or even no picture at all, has any effect upon the *S*s' related imagery scores. Not only does the control group show as much related imagery to the two stimuli as do the corresponding experimental *S*s, but Experimental Group I has about the same amount of related imagery for Stimulus B as does Group II, and Group II has as much related imagery for Stimulus A as does Group I. In addition, it is to be noted that the magnitudes of the related imagery scores for all three groups are comparable to the scores obtained by Shevrin and Luborsky (1958) although the present scores tend to be slightly lower. A somewhat lower related imagery score is to be anticipated in our data since the number of items in our stimulus pictures was somewhat less than in the picture used by Shevrin and Luborsky (21 and 29 items in Pictures A and B as compared with 34 in the Shevrin and Luborsky picture).

Before we conclude that Shevrin and Luborsky's Poetzl phenomenon was solely attributable to an unassessed base rate effect, there is one possibility that needs to be considered. While the two experimental groups do not differ from each other or from the control group on related imagery, it may be that the kind of items appearing in the images differ among the three groups.

We assessed this possibility in two ways. First, we computed the preconscious recall scores for the experimental *S*s by subtracting

their recalled again scores (items in the images that were mentioned in one or more of the attempts at intentional recall) from the related imagery scores. Since the control *S*s had not received a pictorial presentation, no intentional recall score existed for them. However, this problem was solved in the following manner. Each of the 15 control *S*s was randomly paired with one of the 15 *S*s in Experimental Group I and the control *S* was then assigned the intentional recall items of his corresponding experimental mate. The preconscious recall score for a control *S* was computed by subtracting from his related imagery all those items mentioned by his matched experimental *S* during intentional recalls. The control *S*s were then randomly paired with the experimental *S*s in Group II and this procedure repeated. To check upon the randomization of this pairing a second randomization was performed for both groups. Also, parallel computations were again made on data obtained from the perfect agreement and the incomplete agreement conditions. With these measures for the three groups of *S*s we are now in a position to test the possibility that the items comprising the related imagery of the experimental *S*s are different from those of the control *S*s.

If the kind of related imagery items differs among the three groups, then subtracting an experimental *S*'s intentional recall items from a control *S*'s related imagery items should result in a difference of the recalled again scores between the experimental and control groups. The control *S*s should have lower recalled again scores and correspondingly higher preconscious recall scores.

The mean recalled again and preconscious recall scores obtained in this manner are shown in Table 3. As is seen there is little or no difference in the recalled again scores between Group I and the control *S*s or between Group II and the control *S*s. This lack of difference is confirmed by *t* tests run between Group I and the two randomizations of the control group. For the perfect agreement data the values of *t* were .57 for the first randomization and .56 for the second randomization. For the incomplete agreement data the corresponding *t* values were .46 and .18. Similarly comparisons of Group II with the two randomizations of the control group gave *t* values of .43 and .19 for perfect agreement and .86 and 1.02 for incomplete agreement.[2]

TABLE 3
MEAN NUMBER OF ITEMS PRECONSCIOUSLY RECALLED AND RECALLED AGAIN

Group	Perfect Agreement		Incomplete Agreement	
	Preconscious Recall	Recalled Again	Preconscious Recall	Recalled Again
Stimulus A				
Experimental I	3.3	1.6	5.8	4.5
Control First random matching	3.3	1.9	5.9	4.9
Control Second random matching	3.3	1.9	6.4	4.4
Stimulus B				
Experimental II	2.2	1.3	5.3	3.1
Control First random matching	3.0	1.1	6.9	3.9
Control Second random matching	2.9	1.2	6.7	4.1

The failure to obtain significant or even appreciable differences in this comparison renders unlikely the possibility that the kinds of items contained in the related imagery of the control group and experimental groups differ. Since the operation of subtracting an experimental *S*'s intentional recall items from a control *S*'s related imagery leaves both *S*s with similar recalled again and preconscious recall scores, the control *S*'s related imagery must consist in large part of the same items existing in an experimental *S*'s related imagery. In the experimental *S*'s recall, however, they are labeled as recalled again, but the analysis casts considerable doubt as to whether these are items recalled again or just another manifestation of the base rate.[3]

[2] Since the preconscious recall score is obtained by subtracting recalled again items from related imagery there was no point in testing the significance of the means for this measure. Since there was no difference between the groups on related imagery or recalled again scores there could have been no difference in preconscious recall.

[3] Although only four dreams were obtained in each of the experimental groups and two in the control group, identical analyses were carried out on the

One further test of the question of differences in kind of unintentional recall was performed. We computed the number of times each of the 21 items in Stimulus A was mentioned in the imagery of Experimental Group I *S*s and the control *S*s. Eight of the items occurred more frequently in the images of the experimental group but on only three items was the frequency difference 2 or greater (differences of 2, 3, and 7). A similar comparison of the control *S*s with Group II on the 29 items of Stimulus B showed that 6 of the items occurred more frequently in the images of the experimental *S*s, but for only 3 of these items was the frequency difference 2 or greater (differences of 2, 2, and 4).

DISCUSSION

The foregoing results offer no support to the view that unconscious or preconscious perception occurs during tachistoscopic presentation and is subsequently reflected in such indirect measures as daydreams or images. We were unsuccessful in finding any differential influence upon the images of our *S*s as a function of differential prior stimulation. Not only do the control *S*s, who have seen neither stimulus picture, have related imagery scores closely similar to the experimental groups', but the amount of preconscious recall for these control *S*s is also directly comparable (see Table 3). These findings hold true not only when we use the conservative criterion of perfect agreement between independent judges as to the presence or absence of a stimulus item in the *S*'s imagery, but also when we use the more liberal criterion of one or the other judge's considering the item present in the image.

It is possible that our failure to confirm previous interpretations of the Poetzl phenomenon may be due to subtle differences in the population of *S*s used and unrecognizable differences in techniques and methodology. The fact that the magnitude of the related imagery and the preconscious recall obtained in our study is directly comparable to that of Shevrin and Luborsky (1958) would argue against such an interpretation. Since both studies used what would appear to be comparable stimulus material, the finding that any related imagery or preconscious recall in our study was accounted for solely in terms of the base rate would necessarily imply that the Shevrin and Luborsky results were also a function of the unassessed base rate.

The demonstration of the high base rate operating in the use of a classification system like "conceptual similarity," suggests that extreme care should be taken in the evaluation of dreams, images, and other associational data. Such a high rate of coincidental appearance of objects in fantasy may lead one to jump to tenuous conclusions of causality in the interpretation of dreams, projective tests, etc. when, in fact, the relationships perceived are explainable in terms of base rate phenomena due to the remarkable ability of the human mind to see conceptual relations among almost any objects.

We had originally anticipated that the base rate would not account for all of the related imagery and preconscious recall effects and, as was pointed out in the introduction of this paper, we had planned to institute a control for the effect of prior learned associations among objects. Since there was no effect remaining after base rate had been eliminated, the application of this control would have been rather pointless. Nevertheless, it is a control that needs to be considered in evaluating other studies on the Poetzl phenomenon and certainly needs to be taken into account in any future investigations in this area.

It has been pointed out elsewhere (Eriksen, 1960) that the evidence for unconscious perception during a tachistoscopic presentation is, at best highly controversial. Experimenters should therefore be particularly careful in positing unconscious processes as explanations for their data when simpler explanations have not been ruled out.

Even when artifacts such as those discussed here have been excluded there are simpler explanations available for an obtained Poetzl phenomenon that do not require usage of ambiguous concepts such as preconscious perception and primary process thinking. A Poetzl-like effect could be produced as a function of an *S*'s operating characteristic (Goldiamond, 1958). If, as seems reasonable, *S*s tend to adopt a more rigorous criterion for report under the

dreams. The results were comparable to those reported above and there was similarly *no* indication of even a trend in the direction of preconscious recall.

intentional recall condition with its veridical orientation than under the daydream condition, items of which *S* was too unsure to mention in deliberate recall might appear in the less reality oriented daydreams.

SUMMARY

The Poetzl phenomenon was experimentally analyzed using a method similar to that employed by Shevrin and Luborsky (1958). The *S*s in two experimental groups were shown tachistoscopically one of two experimental pictures and were asked to describe it. Daydreams obtained from these *S*s on the following day were analyzed to determine the presence of items conceptually similar to those in the respective pictures. This experiment differed from Shevrin and Luborsky's in using a control group who received no prior picture stimulation but whose images were examined to determine the base rate for conceptually similar items. Also, stricter experimental control was maintained over the instructions to the *S*s.

Analysis of the results provided no support for an interpretation of the Poetzl phenomenon in terms of preconscious perception. Although the number of conceptually similar items appearing in the images that had not been reported in intentional recall was comparable in magnitude to the number reported by Shevrin and Luborsky, it was found that this number did not differ significantly nor appreciably from the base rate obtained from the control group.

Several sources of artifact that occur in experimental designs such as are employed in studying the Poetzl effect were discussed.

REFERENCES

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, **67**, 279–300.

FISHER, C. Dreams and perception: The role of preconscious and primary modes of perception in dream formation. *J. Amer. Psychoanal. Ass.*, 1954, **2**, 389–445.

FISHER, C. Dreams, images, and perception: A study of unconscious-preconscious relationships. *J. Amer. Psychoanal. Ass.*, 1956, **4**, 5–48.

FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. Psychoanal. Ass.*, 1957, **5**, 5–60.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–412.

MALAMUD, W. Dream analysis: Its applications in therapy and research in mental diseases. *Arch. Neurol. Psychiat.*, 1934, **31**, 356–372.

MALAMUD, W., & LINDER, F. E. Dreams and their relationships to recent impressions. *Arch. Neurol. Psychiat.*, 1931, **25**, 1081–1099.

POETZL, O. Experimentell erregte Traumbilder in ihren Bezichungen zum indirekten Sehen. *Zbl. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

SHEVRIN, H., & LUBORKSY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

(Received April 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 504-513

STABILITY OF ACHIEVEMENT AND RECOGNITION SEEKING BEHAVIORS FROM EARLY CHILDHOOD THROUGH ADULTHOOD[1]

HOWARD A. MOSS AND JEROME KAGAN
Fels Research Institute, Yellow Springs, Ohio

THE supposition that selected adult response patterns are established at an early age is a primary assumption of developmental theory. Although literary documents and psychotherapy protocols have provided anecdotal support for this hypothesis, more objective validation has been difficult to obtain. The present paper is a second report that has emerged from a larger project on the stability of childhood behavior. The first paper indicated that dependent behavior in girls showed moderately high stability from the early school years through young adulthood (Kagan & Moss, in press). The present report is concerned with the developmental consistency of two related behaviors: the tendency to strive for (*a*) mastery of selected skills (achievement behavior), and (*b*) social recognition through acquisition of specific goals or behaviors (recognition behavior).

The achievement variable emphasizes mastery of intellectual, athletic, mechanical, and artistic skills as well as competence in specialized crafts. Social recognition is obtained through acquisition of most of the above behaviors. For intellectual competence, athletic ability, acquisition of money, and positions of power in social groups are the primary methods of obtaining social recognition in the cultural milieu of our middle class population. Thus, the overt behaviors involved in achievement and recognition strivings overlap to some degree.

In an attempt to differentiate between these two variables, the investigators evaluated the degree to which the individual's mastery behavior was directed at satisfaction of an internal standard of excellence in order to gain self-approval (achievement motivation), in contrast to seeking approval from the social environment (recognition motivation). This is a difficult differentiation to make. The data to be presented reveal a high, positive correlation between ratings of these two behavioral variables. This interdependence suggests that it may be impossible to measure the "desire to improve at a skill" independent of the individual's "desire for social recognition" for this improvement.

[1] This research was supported, in part, by Research Grant M-1260 from the National Institute of Mental Health, United States Public Health Service.

METHOD

Subjects and General Procedure

The subjects were 36 males and 35 females from the Fels Research Institute's longitudinal population. They were enrolled in the project at birth, during the years 1929–1939. At the time of a recent adult assessment (1957–1959) they were between 20 and 29 years of age. The subjects came from predominantly middle class backgrounds, over half of the group were married; 70% had college degrees or were enrolled in a college, and the majority were living within a 30-mile radius of the institute. The adult group included 55 Protestants, 15 Catholics, and 1 Jew.

The heart of this study consists of correlations between the childhood information on these subjects and their adult behavior. The childhood data included (*a*) longitudinal observations of the child's behavior during the first 14 years of life in a variety of settings, (*b*) observations of the mother-child interaction during these years, (*c*) TAT protocols obtained in adolescence, and (*d*) annual Stanford-Binet intelligence test scores during the ages 5–11. Although the data collected during adulthood (age range 20–29) sampled a variety of techniques, this report utilizes only two sources of adult information, 5 hours of interview, and a TAT protocol.

Longitudinal Observations: Birth to Age 14

As a standard procedure of the Fels longitudinal program, psychologists or psychologically trained personnel summarized their observations of the child in the home, in the Fels nursery school and day camp, and in the subject's public school. The home reports were based on a visit to the home where mother and child were observed for half-day sessions. These home visits were generally made semiannually for the first 6 years of life and annually from 6 to 12. Most of the mothers were interviewed each year for the first 14 years of the child's life. The nursery school summaries were based on semiannual, free-play sessions from age 2.5 to 5. The sessions usually consisted of 15 consecutive half-day periods in groups of 10–12 children. Day camp typically consisted of an annual 2 week session

of half-day periods during age 6–10 in which free and structured group activities were observed. Public school visits, made semiannually, consisted of a half-day observation of the child in his routine classroom activities. Finally, the subjects of age 6–14 were interviewed each year at the institute and a summary of the interview was prepared. All of the longitudinal reports for each subject were collated in chronological order and placed in the subject's individual file.[2]

Scoring of longitudinal variables. A comprehensive list of rating scale variables (seven-point scale) was defined for the purpose of evaluating the narrative material just outlined. The material for each subject was divided into four age periods: 0–3, 3–6, 6–10, and 10–14. The senior author, who had no knowledge of the adult psychological status of the subjects, first read all the information for each subject for age 0–3 and made those ratings for which he had adequate information. Following a period of interpolated work, he studied each subject's material for age 3–6 and again made his ratings. This procedure was repeated for ages 6–10 and 10–14. A period of approximately 6 months intervened between the evaluation of the data for any one subject for each age period. This paper deals only with the stability of achievement and recognition behaviors and abridged definitions of these variables follow.

Childhood Variables

Achievement behavior: (*Rated for ages 0–3, 3–6, and 6–10*). This variable assessed the degree to which the subject tended to persist with challenging tasks, games, and problems, and his involvement in activities in which a standard of excellence was applicable. For 0–3, emphasis was given to persistence with perceptual-motor activities, (e.g., making block towers, stringing beads, drawing, and coloring). For ages 3–6 and 6–10 the greatest weight was given to interest in and persistence with intellectual, mechanical, athletic, and fine motor activities.

For age 10–14 the general achievement variable defined above was differentiated into three variables dealing with different achievement areas (intellectual, mechanical, and athletic).

Intellectual achievement: (*Rated for age 10–14*). This variable assessed the degree to which the subject attempted to master language and numerical skills and showed involvement in the acquisition of knowledge.

Mechanical achievement: (*Rated for age 10–14*). This variable assessed the degree to which the subject attempted to master mechanical skills and manifested involvement in activities such as carpentry, construction of model vehicles, engines and motors, and craft work.

Athletic achievement: (*Rated for age 10–14*). This variable assessed the degree to which the subject attempted to master and showed involvement in athletic activities. These behaviors included swimming, hiking, baseball, football, basketball, tennis, acrobatics, and track events.

Recognition seeking behavior: (*Rated for ages 6–10 and 10–14*). This variable assessed the subject's striving to obtain goals that led to recognition from parents, teachers, and peers. The behaviors emphasized in the rating were (*a*) grades in school and school honors, (*b*) stated desire for status-laden vocations or ostentatious material goods, (*c*) striving for leadership in teams or clubs, (*d*) attempts to get recognition from farm activities (e.g., raise the best calf, the highest corn, etc.).

Maternal Variables

Maternal acceleration of developmental skills in child: (*Rated for ages 0–3, 3–6, and 6–10*). The home visits and maternal interviews yielded information on the mother's behavior and attitudes toward her child. The maternal variable that is directly relevant to the subject's achievement behavior was called *maternal acceleration*. It was defined in terms of the degree to which the mother showed concern over the subject's cognitive and motor development, and the degree to which she exhibited desires for precocious achievement in her child. The rating reflected the degree to which the mother "pushed" the subject's development beyond his abilities and her concern with his general achievement level.

Adult Interview

The junior author, who had no knowledge of the subjects childhood information, interviewed each subject and rated him (seven-point scale) on a variety of variables. The definitions of the variables related to achievement and recognition seeking behaviors follow.

Achievement behavior. This variable evaluated the subject's behavioral attempts to master tasks for which "self-satisfaction" rather than social recognition was the *salient* goal. In achievement behavior, the subject was striving to attain a *self-imposed* standard of excellence. The rating was based on the subject's emphasis and concern with task mastery in his job and avocational pursuits.

Recognition seeking behavior. This variable evaluated the subject's behavioral attempt to obtain symbols of status and social recognition. The rating was based on evidences of strivings for (*a*) vocational recognition, (*b*) academic awards and honors, (*c*) positions of leadership or recognition in community or vocational groups, (*d*) concern with conspicuous material display, (*e*) striving for upward mobility in social class position.

Concern with intellectual competence. This variable assessed the value the subject placed upon intelligence, knowledge, academic achievement, and intellectual superiority regardless of whether the goal was to satisfy inner standards or to obtain social recognition.

Reliability of Longitudinal and Adult Interview Ratings

A random sample of 32 tape recorded adult interviews were independently studied and rated by a second

[2] The staff of the institute during the 24 years of data collection included Alfred L. Baldwin, Thomas W. Richards, Horace Champney, Virginia L. Nelson, assisted by Leah Levinger, Helen Marshall, Mary Frances Hartson, Joan Kalhorn Lasko, Faye Breese, Margaret Slutz, Marjorie Powell, Frances Best, and a group of assistants to whom the authors are indebted. Their efforts made this project possible.

judge to assess the reliability of the junior author's adult ratings. The reliabilities of the longitudinal variables were also assessed through independent ratings, by a second judge, of samples of 50–60 cases at each of the four age periods. The reliabilities of the adult and child ratings were determined by product-moment correlation coefficients.[3] For the adult ratings of achievement behavior, recognition behavior, and intellectual concern the reliability coefficients were .84, .99, and .98, respectively. With the exception of one child behavior variable, the reliabilities of the longitudinal ratings ranged from .74 to .90 with a median coefficient of .81. The one low longitudinal reliability was for child's achievement for age 0–3 ($r = +.35$; $p < .01$; two-tailed).

TAT Achievement Fantasy: Adolescent and Adult Protocols

Early adolescent (median age of 14–6) protocols were available for 67 of the 71 subjects, and all 71 subjects were administered TAT stimuli following the adult interview. The adolescent protocol was based on seven cards from the Murray (1943) series (Cards 1, 5, 14, 17BM, 3BM, 6BM, and 3GF). The male adult protocol was based on 13 cards (4, 8BM, 7BM, 6BM, 12M, 17BM, 13MF, 14, 3BM, 5, 1, 3GF, and 18GF). The adult females were also administered 13 cards (4, 6GF, 12F, 2, 8GF, 17BM, 13MF, 14, 3BM, 5, 1, 3GF, and 18GF). For both the adolescent and adult protocols achievement themes were scored according to the scheme described by McClelland, Atkinson, Clark, and Lowell (1953). Since incidence of the subcategories of the McClelland scoring system were infrequent, only stories in which achievement behavior was the major aspect of the plot were considered. These are scored Ach Th in the McClelland scheme. For the adolescent protocol, there was a lack of comparability among the examiners with respect to the inquiry questions and only the spontaneous verbalization of the subject was scored. Agreement between two independent coders was 95%. The longitudinal and interview ratings of achievement and recognition behavior were made *without knowledge* of the subject's adolescent or adult TAT stories. Thus, the behavior and interview ratings were independent of each other and of the TAT thematic scores.

IQ Change

Each child was given the Stanford-Binet, Forms L and M alternately, annually from ages 5 through 11 by the same psychologist.[4] The mean IQ for the entire Fels population is about 120 (*SD* of 15). For each subject, a smoothed plot of his IQ scores was obtained by averaging his three IQ scores around each age. For example, a child's smoothed or average IQ at age 6 was the result of averaging his IQ scores at ages 5, 6, and 7; his smoothed IQ at age 10 was the average of his IQs at ages 9, 10, and 11. This procedure tends to remove the chance variation associated with any one IQ score and has been used in other studies (Kagan, Sontag, Baker, & Nelson, 1958; Sontag, Baker, & Nelson, 1958). Each subject's smoothed IQ at age 6 was then subtracted from his smoothed IQ at age 10 and the resulting difference was used as a measure of IQ change. As with achievement themes, the child and adult achievement ratings were made without knowledge of the subject's IQ or his IQ change score.

In summary, four independent sources of data were analyzed: child and maternal behaviors for the first 14 years of life, adult behavior, adolescent and adult achievement themes, and childhood IQ change scores.

Statistical Analysis

Relationships among the following variables were evaluated: (*a*) childhood achievement and maternal acceleration ratings with the adult interview ratings, (*b*) adolescent achievement themes with adult achievement themes, (*c*) adolescent and adult achievement themes with the longitudinal and adult ratings, and (*d*) IQ change scores with the childhood and adult ratings. Product-moment correlations were used except when the TAT achievement score was involved. Since achievement themes were not normally distributed, contingency coefficients[5] were used for all tests of association using this variable. Mechanical achievement for age 10–14 was the only variable for which there was a significant sex difference; the boys having a higher mean rating than the girls ($p < .05$; two-tailed).

RESULTS

Stability of Achievement and Recognition Behaviors

Table 1 presents the relationships between the child and adult ratings of achievement and recognition behavior, as well as the relation between maternal acceleration and the adult achievement variables. There are several important results in this table. The rating of achievement behavior for age 6–10 showed a significant, positive association with all three adult variables for both sexes. The rating of achievement for age 3–6 was predictive of adult behavior for the females but not for the males, a finding that suggests the earlier emergence of stable achievement strivings in girls' development than in boys. Of the three achievement behaviors rated for age 10–14, only intellectual mastery was predictive of adult achievement for both sexes. Involvement in mechanical activities was predictive of adult achievement for boys but not for girls. Athletic achievement showed no relationship to the

[3] All correlations were corrected for restricted range of scores using a procedure described by Peters and Van Voorhis (1940, pp. 395–398).

[4] Virginia L. Nelson administered all of the IQ tests.

[5] The contingency coefficients were based on chi squares computed from Mood's likelihood ratio test for a 3 × 2 distribution (Mood, 1950, p. 257).

TABLE 1
RELATION BETWEEN LONGITUDINAL RATINGS OF CHILDHOOD ACHIEVEMENT AND EARLY MATERNAL ACCELERATION WITH ADULT ACHIEVEMENT BEHAVIOR
(product-moment correlations)

Childhood Variables	Age	Adult Variables					
		Achievement		Recognition		Intellectual Concerns	
		Males	Females	Males	Females	Males	Females
Recognition	6–10	.47***	.40**	.42**	.48***	.37**	.55****
	10–14	.25	.20	.36*	.39**	.24	.40**
Achievement	0–3	−.12	−.02	.01	−.22	−.08	−.02
	3–6	−.03	.45**	−.11	.49***	.13	.44**
	6–10	.46***	.38**	.57****	.51***	.69****	.49***
Achievement							
Intellectual	10–14	.40**	.42**	.60****	.56***	.66****	.49***
Mechanical	10–14	.20	.20	.46**	.02	.47**	.27
Athletic	10–14	−.18	.01	−.17	−.09	−.47**	.02
Maternal Acceleration	0–3	.22	.36*	.44*	.41*	.09	.36*
	3–6	.31	.09	.24	.12	.42**	.12
	6–10	.14	.33*	.16	.23	.32*	.43**

* $p < .05$; one-tailed.
** $p < .02$; one-tailed.
*** $p < .01$; one-tailed.
**** $p < .001$; one-tailed.

rating of general adult achievement, and was negatively associated with intellectual concern for adult males ($p < .02$).

Recognition seeking behavior for age 6–10 was also predictive of adult achievement behavior. A few of the child variables were moderately intercorrelated and the three adult variables were highly intercorrelated (Tables 4, 5, and 6). This lack of independence makes some of the stability correlations between childhood and adulthood somewhat redundant.

Maternal Acceleration and Adult Behavior

Maternal concern with the child's developmental progress during the first 10 years of life showed low to moderate correlations with adult achievement behavior. The maternal rating for age 6–10 was not a better predictor of adult behavior than the maternal rating for the first 3 years of life. Moreover, the age 0–3 rating was associated with all three adult, achievement variables for girls, while it predicted only recognition behavior for adult males.

Stability of TAT Achievement Fantasy

Although different sets of TAT pictures were used in obtaining the adolescent and adult protocols, the three pictures that usually elicited achievement stories were presented at both administrations. Cards 1, 14, and 17BM, which elicited 77% of all the achievement themes, were common to both protocols. The strong tendency for these particular cards to elicit achievement themes has been noted in another study (Kagan & Moss, 1959). A typical achievement theme to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist. A typical achievement story to Card 17BM involved a person who was in a rope climbing contest and wanted to do his best to win. A common achievement story to Card 14 concerned an artist or student who had been working hard and was looking forward to fame and success as a result of his accomplishments.

The stability of the TAT achievement score between the adolescent and adult protocols was determined through the use of contingency coefficients. The stability coefficients were .34, .36, and .31 for boys, girls, and total group ($p < .10$, $< .05$, $< .02$; one-tailed). Thus, achievement themes also showed some degree of stability over this 10-year period. These data extend the findings of an earlier investigation (Kagan & Moss, 1959), in which the authors reported a 3 year stability coefficient of .32 ($p < .01$) for achievement themes obtained at median ages of 8–9 and 11–6. The stability coefficients between the adolescent and adult protocols are of the same magnitude as those found for the earlier age period.

TABLE 2

RELATION BETWEEN TAT ACHIEVEMENT THEMES AND CHILD AND ADULT ACHIEVEMENT BEHAVIOR
(contingency coefficients)

Longitudinal Variable	Age	Adolescent TAT (Median age 14-6)		Adult TAT (Median age 25)	
		Males	Females	Males	Females
Recognition	6–10	.21	.17	.26	.22
	10–14	−.20	−.18	−.31	.39
Achievement	0–3	−.15	−.25	−.20	.16
	3–6	.42*	.19	.19	.36
	6–10	.24	.15	.13	.30
Achievement					
Intellectual	10–14	.30	−.25	.26	.16
Mechanical	10–14	.31	−.62***	.63****	−.50**
Athletic	10–14	−.20	.12	.12	.17
Maternal acceleration	0–3	.11	.51**	.25	−.08
	3–6	.37	.23	.27	.28
	6–10	.51***	.26	.24	.41*
Adult interview variables (median age 25)					
Recognition		.17	.25	.40**	.52****
Achievement		.19	.44***	.37*	.52****
Intellectual Concerns		.44***	.25	.31	.59****

* $p < .05$; one-tailed.
** $p < .02$; one-tailed.
*** $p < .01$; one-tailed.
**** $p < .001$; one-tailed.

Validity of Achievement Themes: Relations with Child and Adult Behavior

Contingency coefficients were computed relating the occurrence of adolescent and adult achievement themes with the longitudinal and adult achievement ratings. These results are presented in Table 2. The highest and most consistent relations were between the adult achievement themes and adult interview ratings. The only significant relation between adult themes and the childhood ratings held for mechanical achievement ($C = +.63$; $p < .001$ for boys, and $-.50$; $p < .02$ for girls).

The adolescent TAT was also more predictive of adult behavior than it was of the childhood ratings. Adolescent achievement themes predicted adult achievement behavior for women ($C = +.44$; $p < .01$) and intellectual concerns for men ($C = +.44$; $p < .01$). Adolescent achievement themes showed minimal association with the child's achievement behavior. The only significant positive association was with age 3–6 achievement for boys. Once again the rating of mechanical achievement for girls was negatively associated with achievement themes. This negative correlation may be due to the fact that this is the only variable for which markedly different behavioral referents were used in rating the two sexes. For boys, involvement in carpentry, engines, motors, and model airplanes was emphasized in the rating. These activities are sex-typed and girls showed no interest in them. Participation in craft work (making jewelry, leather articles) and sewing was also used as evidence of involvement in mechanical activities and girls tended to choose these behaviors.

Maternal acceleration during the first 10 years of life showed suggestive relationships with the adolescent achievement themes. For example, maternal acceleration for age 0–3 predicted achievement themes at adolescence for girls ($C = +.51$; $p < .02$), but not for boys. Maternal acceleration for age 6–10 predicted adolescent achievement themes for boys ($C = +.51$; $p < .01$) and adult achievement themes for girls ($C = +.41$; $p < .05$).

In summary, the adult and adolescent TAT stories showed moderate correlations with adult achievement but minimal association with the childhood achievement ratings. Maternal acceleration was associated, to some degree, with adolescent achievement themes.

IQ Increase and Achievement Behavior

The difference between the child's smoothed IQ at age 6 and 10 was used as a measure of IQ

TABLE 3

RELATION BETWEEN IQ CHANGE AND CHILDHOOD AND ADULT ACHIEVEMENT VARIABLES
(product-moment correlations)

Longitudinal Variables	Age	Males	Females
Recognition	6–10	.24	.21
Recognition	10–14	.41**	.09
Achievement	0–3	.13	.04
Achievement	3–6	−.02	.24
Achievement	6–10	.39***	.47***
Achievement-intellectual	10–14	.37*	.41**
Achievement-mechanical	10–14	.15	.14
Achievement-athletic	10–14	−.16	−.46**
Maternal acceleration	0–3	−.06	.20
Maternal acceleration	3–6	−.03	−.12
Maternal acceleration	6–10	.10	.54***
Adult interview variables			
Recognition		.48***	.25
Achievement		.38***	.38**
Intellectual concern		.49***	.42***

* $p < .05$; one-tailed.
** $p < .02$; one-tailed.
*** $p < .01$; one-tailed.

change. Earlier studies (Kagan et al., 1958; Sontag et al., 1958) have demonstrated that the amount of increase in IQ correlated both with independent behavioral indices of achievement strivings for age 6–10, and with early adolescent (age 10–14) achievement stories. These latter relations remain significant when the influence of the child's IQ at age 6 is statistically controlled. The present data allowed for a partial validation of these results and an extension of the Sontag et al. findings for adolescence and adulthood.[6]

Table 3 presents the correlations between changes in IQ during age 6–10 and the longitudinal and adult behaviors. The amount of IQ increase was a fairly sensitive predictor of both intellectual achievement for age 10–14 ($r = .37$ and .41 for boys and girls; $p < .01$), and concern with intellectual competence in adulthood ($r = .49$ and .42; $p < .01$). These results support and extend the earlier studies and indicate that amount of IQ increase during the first 4 years of school is a moderately accurate index of the subject's motivation to master intellectual tasks during adolescence and early adulthood. It is important to note that IQ change showed no relation to mechanical or athletic strivings for boys, and was negatively associated with athletic achievement for girls ($r = -.46$; $p < .02$). Thus, IQ increase is not a general measure of achievement strivings for all areas of task mastery. The IQ change measure predicts all three adult achievement ratings because the three adult variables are heavily weighted with concern over intellectual competence. Finally, the maternal acceleration rating for age 6–10 showed a positive relation with IQ change for girls ($r = +.54$; $p < .01$) but not for boys.

[6] In the present sample of 71 subjects, 50% of the males and 20% of the females overlapped with the group of 70 subjects studied by Sontag, Baker, and Nelson (1958). However, their data only dealt with the period from 3 to 10 years of age. The present behavioral material covered adolescence and early adulthood.

TABLE 4

INTERCORRELATIONS AMONG VARIABLES RATED FOR AGE 6–10

	Recognition	Achievement	Maternal Acceleration
Recognition	—	.77***	.57***
Achievement	.60***	—	.59***
Maternal acceleration	.39*	.44***	—

Note.—Data for males are in upper right; for females, in lower left.
* $p < .05$; two-tailed.
*** $p < .01$; two-tailed.

Intercorrelations among the Measures

There were, as might be anticipated, positive correlations among the achievement and recognition ratings. Tables 4, 5, and 6 present the intercorrelations among the variables that were rated for ages 6–10, 10–14, and adulthood. The correlations for the males are above and to the right of the diagonal; the female data are to the left and below the diagonal.

For all three age periods there were high, positive correlations among the achievement and recognition variables. For age 10–14, recognition behavior was highly correlated with achievement strivings in the intellectual area, but only minimally related to mechanical or athletic achievement. This finding suggests that, for this middle class sample, mastery of intellectual skills is the primary method chosen to obtain social recognition. Perhaps for lower class samples this generalization might be less valid. The high correlations between recognition behavior and intellectual concern in adult-

TABLE 5
INTERCORRELATIONS AMONG VARIABLES RATED FOR AGE 10–14

	Recognition	Intellectual Achievement	Mechanical Achievement	Athletic Achievement
Recognition	—	.74***	.23	.04
Intellectual achievement	.60***	—	.53***	−.11
Mechanical achievement	.17	.32	—	−.23
Athletic achievement	.24	.12	.07	—

Note.—Data for males are in upper right; for females, in lower left.
*** $p < .01$; two-tailed.

TABLE 6
INTERCORRELATIONS AMONG ADULT INTERVIEW VARIABLES

	Achievement	Recognition	Intellectual
Achievement	—	.72***	.73***
Recognition	.79***	—	.72***
Intellectual	.77***	.84***	—

Note.—Data for males are in upper right; for females, in lower left.
*** $p < .01$; two-tailed.

hood, together with the fact that maternal acceleration predicted both variables, suggests that it is difficult to separate "recognition seeking behavior" from "attempts to improve intellectual competence."

DISCUSSION

Stability of Achievement Strivings

The results indicate that strivings for intellectual mastery are moderately stable from the school years through early adulthood. This behavioral disposition emerges as a stable phenomenon at ages 3–6 for girls and 6–10 for boys. The stability of the behavior ratings is paralleled by the moderate stability of TAT achievement stories over a shorter age span. Moreover, achievement stories in adolescence and adulthood also predicted the adult behavior ratings. This consistent cluster of correlations adds construct validity to the TAT achievement variable and support to the conclusion that this class of behaviors is stable over time.

Involvement in athletics for age 10–14 showed no strong, positive relation to either IQ increase or adult achievement behavior and, in a few instances, negative relationships occurred. This was not because the interviewer failed to assess adult involvement in this particular activity. Rather, many of the adults who had been involved in athletics as early adolescents were not overly concerned with task mastery as adults and they tended to avoid intellectual activities.

The majority of the sample regarded positions of responsibility, intellectual challenge, and knowledge of the environment as highly desirable goals. If a subject had strong achievement motives he tended to gratify them through intellectually oriented endeavors. It is suggested that the mass media and social environment differentially emphasize the importance of different skills in accordance with the sex and age role characteristics of the individual. For adults, there tends to be an emphasis on intellectual competence and a de-emphasis on active mastery of athletic skills. Moreover, intellectual mastery is less involved in potential sex role conflict than mechanical or athletic behaviors. To excel at sports is one of the defining characteristics of masculinity. Some boys become involved in athletics in order to maintain their sex role identity and avoid peer rejection. An athletic girl will be subject to peer rejection for excessive participation in athletics. Thus, athletic mastery is under the control of motives and conflicts related to sex role identification in addition to needs for task mastery.

This latter point raises the question of the appropriate definition of achievement behavior and motivation. It is suggested that the concept of a general achievement motive is too broad a term, and it may be useful to replace this construct with a series of variables that relate to more specific behaviors. It seems more reasonable to talk about "desire to improve intellectual skills," or "desire to improve athletic skills" than to use the more global concept of need achievement. Individuals strive to perfect skills in different areas, and the motivations for these strivings are multiple. Prediction and comprehension of these phenomena might be facilitated if there was some differentiation among the behaviors and motives that are involved in task mastery.

The lack of predictive power of age 0–3 mas-

tery behavior might have been due to the greater difficulty in rating this variable (the interrater reliability was .35). On the other hand, the behavioral referents for this rating differed from those used to assess mastery for the older age periods. Since 2-year-olds do not initiate intellectual or athletic mastery behavior, persistence with simple, perceptual-motor tasks (stringing beads, building towers) was the basis for this early rating. A high rating for 0–3 reflected a high threshold for satiation with simple, sensorimotor activities. At the older ages, the achievement rating was based on involvement with problem solving behaviors that were more similar in form to adult achievement behavior. The age 0–3 rating is dynamically different from the symbolic behaviors that characterize achievement during the preschool and school years.[7] This statement is supported by the fact that achievement for age 0–3 was negatively correlated with achievement for age 3–6 ($r = -.20$) and age 6–10 ($r = -.03$), and showed no relationship to achievement themes or IQ change. Persistence with simple sensorimotor tasks during the first 2 or 3 years of life is not an index of future intellectual, achievement strivings. The 2-year-old who will sit for 20 minutes trying to put a peg in a hole is not necessarily the ambitious scholar of the fifth grade.

Maternal Acceleration and Achievement

The ratings of maternal concern with the child's developmental skills were heavily weighted with encouragement of intellectual progress. The most consistent correlates of maternal acceleration were found with the ratings of adult concern with intellectual competence. Maternal acceleration for age 0–3 was slightly more predictive of adult behavior for girls than for boys. Similarly, maternal acceleration for age 6–10 was more predictive of IQ increase for daughters than for sons. The sex difference between these latter two correlations was significant at the .05 level. It is suggested that since the girl was more likely than the boy to identify with the mother, maternal encouragement of intellectual mastery should have had a greater effect on the development of the girl than on the boy.[8]

TAT Achievement Stories

Achievement themes on the TAT were moderately stable and were correlated with adult achievement behavior. The fact that the correlations were as high for females as for males, although the three critical cards illustrated male heroes, raises some question concerning the validity of the hero hypothesis. Since Cards 1, 14, and 17BM all picture a male in a potential achievement situation, one might expect that achievement themes for women would not be highly correlated with their achievement behavior. The present results indicate that the production of achievement themes may be more influenced by the subject's conception of what behaviors are appropriate for the hero, than by the degree of identification of storyteller with hero. Perhaps high achievement girls conceptualize the male role as being more associated with task mastery than do low achievement girls.

Atkinson (1957) has suggested that achievement themes have differential validity depending on whether or not the criterion task engages the subject's motivation. The achievement variables used in this study (ratings of overt behavior, IQ increase scores) measured "real life" behaviors that would be expected to engage the subject's motivation. The positive correlations obtained indicate that achievement themes are valid indices of intellective mastery when the conditions under which the behavioral samples are obtained are motive arousing.

Limitations on Generalizability

Although the stability correlations for achievement behavior are fairly high, the nature of this particular sample favored stability. The social milieu of these subjects remain constant throughout the first 17 years of their lives, and the parents and peers of these subjects retained their same values. The degree of stability obtained with this sample might not

[7] The differences in the content of mastery behaviors for ages 0–3 and 3–6 are analogous to Piaget's (1952) description of intellectual development. Piaget suggests that during the first 2 years the child is in the sensorimotor stage of intelligence in which simple perceptual-motor activity is salient. During the subsequent preschool years the child's intellectual activity becomes more symbolic and more comparable to adult problem solving behavior.

[8] Research in progress at the institute, under the direction of Vaughn J. Crandall and Walter Katkovsky is assessing the role of both mother and father in the adoption of achievement motives and behaviors.

hold for populations that were more mobile, for different ethnic or social class groups, or for children subjected to major developmental traumata.

Social Reinforcement and Stability

The stability of achievement behavior is congruent with general reinforcement theory. Each time achievement strivings are rewarded through social approval or internal feelings of satisfaction, the strength of this behavioral tendency should be increased. If achievement strivings lead to failure, these behaviors should extinguish. The child who attains scholastic honors through effort is rewarded by the social environment, and this experience frequently leads to an expectancy of future success for similar behavior. This rewarding experience, coupled with the strong cultural approval for intellectual competence, increases the probability that the child will continue to engage in intellectual tasks. On the other hand, persistent failures in intellectually challenging situations are likely to lead to an expectancy of failure, and these expectancies can result in avoidance and/or withdrawal from involvement in intellectual behavior.

SUMMARY

This paper summarized results from a larger investigation on the stability of behavior from childhood through adulthood. This investigation dealt specifically with the long term stability of achievement and recognition seeking behaviors in subjects who were part of the Fels Research Institute's longitudinal population.

The subjects were 36 males and 35 females for whom extensive longitudinal information was available from birth through 14 years of age. One psychologist, who had no knowledge of the adult behavior of these subjects, studied narrative reports based on observations of the child in a variety of settings, and rated each child on achievement and recognition seeking behaviors for four age periods: 0–3, 3–6, 6–10, and 10–14. In addition, ratings were made for maternal acceleration of developmental skills for the first three age periods. A second psychologist, who had no knowledge of the childhood information, interviewed each subject in adulthood (age range 20–29) and rated him on three variables related to achievement and recognition seeking behavior in adulthood.

In addition, the following information was available for most subjects: (*a*) a 7 card TAT protocol administered during early adolescence (median age 14–6) and a 13 card TAT protocol following the adult interview, and (*b*) annual Stanford-Binet IQ tests from ages 5 through 11 which furnished an IQ change score over the years 6–10.

The major results were as follows:

1. Both achievement and recognition striving behaviors for age 6–10 showed significant positive correlations with similar behaviors during adulthood, the correlations ranged from .38 to .68 and all were significant at the .05 level or better.

2. Involvement in intellectual activities for age 10–14 showed high positive correlations with achievement behavior in adulthood while involvement in athletics during these years showed no positive relationship with adult achievement behavior. Involvement in mechanical tasks for age 10–14 showed positive correlations with adult achievement behavior for boys but not for girls.

3. Maternal acceleration of the child's developmental skills during the first 3 years of life predicted adult achievement behavior for women but not for men. Maternal acceleration of developmental skills during age 6–10 showed moderate correlations with adult concern with intellectual competence in both sexes.

4. Achievement stories told during adolescence and adulthood showed high positive correlations with adult achievement behavior. Maternal acceleration of developmental skills showed suggestive correlations with the occurrence of achievement stories in both adolescence and adulthood.

5. The amount of increase in IQ score during the years 6–10 showed high positive correlations with the ratings of achievement behavior during adulthood. In addition, increase in IQ showed positive correlations with age 6–10 achievement behavior, and with strivings for intellectual competence during age 10–14. Amount of IQ increase showed negative correlations with athletic achievement for age 10–14.

The results suggested that achievement strivings during the first 4 years of school are a

moderately good index of future achievement behavior during adolescence and adulthood. There were high correlations between strivings for social recognition and intellectual achievement and it was suggested that these variables are intimately related in a middle class population.

REFERENCES

ATKINSON, J. W. Motivational determinants of risk taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

KAGAN, J., & MOSS, H. A. The stability and validity of achievement fantasy. *J. abnorm. soc. Psychol.*, 1959, **58**, 357–364.

KAGAN, J., & MOSS, H. A. The stability of passive and dependent behavior from childhood through adulthood. *Child Develpm.*, in press.

KAGAN, J., SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality and IQ change. *J. abnorm. soc. Psychol.*, 1958, **26**, 261–266.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.

MURRAY, H. A. *Thematic Apperception Test Manual*. Cambridge: Harvard Univer. Press, 1943.

PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.

PIAGET, J. *The origins of intelligence in children*. New York: International Univer. Press, 1952.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, 1958, **23**, No. 68.

(Received April 11, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 514-519

A REINTERPRETATION OF SOME PATHOLOGICAL DISTURBANCES IN CONCEPTUAL BREADTH[1]

LOREN J. CHAPMAN[2]
University of Kentucky

MANY clinicians and experimental investigators have reported that schizophrenics err in their assignment of objects to common conceptual categories, but the various writers differ on the nature of the error. Some writers have described it as an inappropriate enlarging of the application of a conceptual category, while others have described it as an inappropriate restricting. Observers usually infer inappropriate enlarging when subjects overinclude, i.e., include incorrect objects in a conceptual category, and they infer inappropriate restricting when subjects overexclude, i.e., exclude from a category objects that belong in it. Cameron (1939) emphasizes the first of these, but it should be pointed out that he uses the term "overinclusion" to refer to a wider range of phenomena. Writers who regard schizophrenic thinking disorder as "concrete," such as Goldstein and Scheerer (1941), put primary emphasis on overexclusion. Using more specialized tests, Chapman and Taylor (1957) and Epstein (1953) found only overinclusion, not overexclusion. Zaslow (1950) claimed to find both, but as Kugelmass and Foundeur (1955) pointed out, the results are doubtful because of a lack of reliability of Zaslow's one item test. The present investigation is an attempt to reconcile the apparently conflicting observations of these writers.

Schizophrenics often appear to deal with test materials in ways which they prefer rather than the ways in which they are instructed. It may be that the more basic phenomenon in errors of conceptual sorting is, at least in part, not a tendency toward overexclusion or overinclusion, but a tendency to use concepts of a specific preferred breadth regardless of appropriateness. Thus, if a schizophrenic is instructed to sort by a more narrow concept than he prefers, he will be more overinclusive; but if he is asked to sort by a broader concept than he prefers, he will be more overexclusive. In other words, what may in part underlie the frequently observed inappropriate enlarging or restricting of the application of common concepts is in fact a preference for concepts of a particular breadth.

The present study is an extension of a method partially developed earlier (Chapman & Taylor, 1957) when schizophrenics were compared with normal subjects on a card sorting test designed to measure inappropriate enlargement or restriction of the application of common conceptual categories. Each subject was given a pile of cards, each of which had printed on it the name of an object, and he was asked to sort all those that fell into a named conceptual category into one pile (e.g., all the fruit) and everything else into a second pile. The finding was that the schizophrenics committed more errors of overinclusion than the normals. Moreover, items which were more often incorrectly included were ones which shared with the correct objects another common conceptual category which subsumed the classification specified in the instructions. These results suggested that the subject's overinclusion errors resulted from his sorting by broader concepts than he was instructed to use. However, the schizophrenics did not commit more errors of overexclusion than the normal subjects.

In the present study the breadth of the conceptual category is varied. The hypothesis is that as the conceptual category specified in the instructions becomes broader, schizo-

[1] This study was supported by a research grant (M-1416) from the National Institute of Mental Health, United States Public Health Service.

[2] The author is indebted to Alvin Burstein and Dorothy Day for assistance in testing the subjects and computing the statistics, and to Jean P. Chapman for her helpful suggestions in the preparation of the manuscript. He also wishes to thank J. Herbert Maltz of Chicago State Hospital and Alexander Harvey of Downey Veterans Administration Hospital for assistance in obtaining schizophrenic patients, Gabriel Ferrazzano of the United States Public Health Service Hospital in Chicago, both J. F. Vanerio and Pearl Baker of the Illinois Central Hospital for assistance in obtaining normal subjects, and Roy Brener of Hines Veterans Administration Hospital for assistance in obtaining the brain damaged patients.

phrenics increase their errors of overexclusion but decrease their errors of overinclusion.

Since brain damaged patients have also been reported to show a disorder in application of conceptual categories, it seemed desirable to test a group of such patients as well as a schizophrenic and normal group in order to compare the two patient populations on overinclusive and overexclusive behavior. No specific prediction was made concerning the effect of varying the breadth of conceptual categories on overinclusion and overexclusion by brain damaged patients.

METHOD

Two sets of tasks were used, one designed primarily to elicit errors of overinclusion, the other to elicit errors of overexclusion.

Tasks primarily eliciting overinclusion. The subject was given a pile of 30 cards, each of which named one object, the objects falling into three categories, e.g., 10 fruit, 10 vegetables, and 10 items of sports equipment. The experimenter presented an instruction card naming one of the categories. This card remained visible throughout the sorting. The subject was asked to sort the cards into two boxes: e.g., fruit into one box and everything else into another. The measure of overinclusion was the number of incorrect cards included with the fruit. Errors of overinclusion tend to be elicited on this task because such errors are often the result of sorting by a broader concept than the one specified in the instructions. In the example, the broader concept of edibles would include vegetables as well as fruit. In addition to this measure of overinclusion, the task provided a measure of the extent to which overinclusion occurs on the basis of the subject's using excessively broad concepts. If the subject sorted by a broad concept of a similar kind that subsumed the one specified by the instructions, he more often included items from the incorrect but similar category (e.g., vegetables) than items from the dissimilar category (sports equipment). However, errors of other kinds—those due to carelessness, cards sticking together, inability or unwillingness to take the test, or sorting by another unidentified concept—would result in placements in the two boxes equally often. This does not imply, of course, that the broader concept offered in the task is the only broader concept toward which the patient is inclined. Overinclusive tendencies on the basis of similarity would be expected to lead toward many kinds of responses other than the ones measured here. Thus, a score of number of vegetables placed with the fruit, minus the number of sports equipment that are so placed, measured the tendency to sort by inappropriately broad concepts, independent of errors of other types. In order that differences in sorting the three categories of objects would not be attributable to differences in their familiarity, the three sets of words were equated on frequency of occurrence in the language as indicated by Thorndike and Lorge's (1944) average of three word counts.

The "fruit" concept was used as a comparatively narrow concept, and the 30 cards used for it are called here Task IA-Incl. (Level I). A second task with a comparatively narrow concept was Task IB-Incl. for which the instructed concept was "birds." In addition to the names of 10 birds, the subject was given 10 insect names (incorrect similar) and 10 words referring to articles of clothing (incorrect dissimilar).

For sorting by broader concepts (called here Level II concepts), the materials from Level I tasks were combined with additional cards as described below. For Task IIA-Incl., the 30 cards of IA-Incl. (fruits, vegetables, and sports equipment) were combined with 10 cards naming trees, 10 naming flowers, and 10 naming musical instruments, with the instructed concept being "Things that a person can eat." The 20 fruits and vegetables were scored as correct, the 20 trees and flowers as incorrect similar, and the 20 musical instruments as incorrect dissimilar. For Task IIB-Incl., the 30 cards from IB-Incl. (birds, insects, and clothing) were combined with 10 cards naming wild animals, 10 domesticated animals, and 10 tools. The subject was instructed to sort out the "Creatures that fly." The 20 birds and insects were scored as correct, the 20 animals as incorrect similar, and the 20 tools and clothing as incorrect dissimilar.

It is seen that on Task IIA-Incl. as well as IIB-Incl. the instructed concept was by definition broader than the corresponding one at Level I, because in each case the Level II concept subsumed Level I as a subconcept. The hypothesis was that on Level II there would be fewer errors of overinclusion than on Level I.

Tasks eliciting overexclusion. In order to elicit errors of overexclusion, the same combinations of cards were used as in the tasks testing for errors of overinclusion, but subject was instructed to sort by different concepts. For example, for Task IA-Excl., the concept was "Things that a person can eat," and for Task IB-Excl., it was "Creatures that fly." The concept for IIB-Excl. was "Things that can move without outside help," and for IIA-Excl., "Things that grow." These concepts subsume more of the available items and hence provide more opportunity for overexclusion errors on that particular pile of cards than do the instructions designed to elicit overinclusion. Also, the expectation was that if a subject were disposed toward errors of overexclusion, he would have a tendency to sort by more narrow concepts than specified in the instructions. Because the conceptual category the subject was asked to sort by always included as subconcepts two common and readily identifiable categories of objects, these tasks were expected to elicit overexclusion errors. In addition, the concepts used for both Level I and II on these tasks are one level of abstraction higher than those in the tasks eliciting overinclusion because the instructed concepts subsume as a subclass the concepts used with the same response cards in eliciting overinclusion.

Additional tasks. In addition to these tasks, four others were given, which, although they do not fit into the scheme of levels of abstraction, give additional data on tendencies toward overinclusion and overexclusion. Task IC-Incl. consisted of 10 domesticated animals, 10 wild animals, and 10 tools, with the instruction "Animals that are kept on the home or on the farm."

Task IC-Excl. used the same cards with the instruction "Animals."

Tasks ID-Incl. consisted of 10 trees, 10 flowers, and 10 musical instruments, with the instruction "Names of trees." Task ID-Excl. used these cards with the instruction "Things that grow."

There were 12 separate sortings in all. Every subject received all 12 in a prearranged order such that the subject was never asked to do consecutive sortings by the same conceptual category, and the overinclusion and overexclusion tasks as well as tasks from the two levels were intermixed. This order was established by a circular plan, with consecutive subjects from a given experimental group beginning the experiment with a sorting which was one removed from that of the previous subject. This design was intended to minimize the possibility that differences between groups of subjects as to their performance on the various sortings were due to practice or fatigue effects.

Subjects. Subjects were white males who showed no indication of mental deficiency and who had completed at least the sixth grade. They included 52 schizophrenics, 39 brain damaged patients, and 33 normal subjects.

The schizophrenic subjects were chronic patients of whom 40 were on a "regressed" ward at Chicago State Hospital. Seven were from other Chicago State Hospital wards, and 5 were from Downey Veterans Administration Hospital. Also, none of the patients had received recent somatic therapy, and none was being treated with psychopharmacological agents.

Although no attempt was made to select patients by diagnostic subtype, the schizophrenic group was found to include 6 diagnosed as paranoid, 16 hebephrenic, 6 catatonic, 2 simple, and 22 without classic subtype. No patient was included for whom there was any suspicion of brain damage. All patients had spent at least one year in the institution prior to testing, and most of them had spent a large part of their lives in the hospital.

The brain damaged patients were obtained at Hines Veterans Administration Hospital. Only those patients were included who were diagnosed as suffering from cerebral damage. Three subjects bore diagnoses of brain tumor, 3 of head injuries, 5 of multiple sclerosis, 1 of Huntington's chorea, 1 of chronic brain syndrome, 24 of cerebral vascular disorders, and 2 others. No brain damaged patient was included who appeared on the basis of medical history or interview to be suffering from severe emotional disorder. The brain damaged patients as a group had been hospitalized for a shorter term than the schizophrenics. About one-third of them had entered during the previous 3 months, and none had been hospitalized longer than 5 years. Since these patients were being tested in order to study the nature of cognitive disturbance resulting from brain injury, only those patients were included who appeared by clinical interview to have some intellectual loss or disturbance. This decision was usually made on the basis of the patient's difficulty in communication, serious disturbance of memory, or severe lack of alertness.

The normal subjects were ambulatory, psychiatrically normal medical and surgical patients at the United States Public Health Service hospital and the Illinois Central Hospital in Chicago. The mean age of the three groups was schizophrenic 44, brain damaged 44, and normals 45. Mean education was 10.2, 10.6, and 10.1, respectively. The three groups did not differ statistically with respect to mean age or education. However, the variances of both age and education were not homogeneous, the values being significantly greater for the brain damaged patients than for either of the other two groups ($p = .05$ or less). The variances did not differ significantly between the schizophrenic and normal groups.

RESULTS

Although the overinclusion tasks were designed to elicit overinclusion errors, some overexclusion errors could occur. Similarly, overinclusion errors could occur on the overexclusion tasks. Table 1 lists mean total overinclusion and overexclusion error scores at each level for both sets of tasks as well as probability statements concerning differences in error scores between levels. Since there were only half as many cards and therefore half as many opportunities for both kinds of error on Level I as on Level II, the scores on Level I were doubled to make them comparable. Because the distributions were extremely skewed and the groups small, all differences within groups were tested for significance by a two-tailed Wilcoxon test, and all differences between groups by a two-tailed Mann-Whitney test.

As seen in Table 1, the schizophrenic patients, as predicted, showed changes in number of errors between Level I and Level II that were in opposite directions for the two kinds of errors. Overexclusion errors increased from Level I to Level II, and overinclusion errors decreased. Both differences were significant. The brain damaged patients also showed a significant rise in overexclusion errors between task levels, although the drop in overinclusion errors was significant only on the overinclusion tasks, not on the combination of both sets of tasks ($p = .18$). For the normal subjects, the differences are in the same direction but not significant.

Let us now consider the question of whether the overinclusion errors resulted from use of excessively broad concepts. Table 2 breaks down the overinclusion errors on the overinclusion tasks into those that occurred on items similar to the conceptual category specified for sorting as against those dissimilar. As in the Chapman and Taylor study (1957), the incorrect inclusions were more often from the

similar category than from the dissimilar; whereas random errors, or errors other than those resulting from sorting by an excessively broad concept, fell equally often in similar and dissimilar errors. For example, when instructed to sort by "fruit," subjects included vegetables more often than sports equipment. This indicates that the errors of overinclusion resulted from sorting by an excessively broad concept, since the incorrect similar items together with the correct items formed a slightly broader conceptual category than the one the subject was asked to sort by. The data from the two additional tasks yielded comparable results.

Of more interest to the present study, however, is the finding that the difference in overinclusion behavior between Levels I and II is primarily attributable to a difference in the incorrect similar category. A score of similar errors significantly distinguishes the two levels at the .01 level for both the schizophrenics and the brain damaged, while, as expected, the score of dissimilar errors does not distinguish the levels for either group. These findings were investigated more directly by computing for each subject a score of similar errors minus dissimilar errors. As discussed above, this difference score is a relatively pure measure of sorting by excessively broad concepts with other kinds of errors controlled. On this score, the two levels were distinguished for both pathological groups ($p < .001$). None of these differences was significant for the normal group. We may conclude, therefore, that for the patients there is a difference between the two levels not only in overinclusion errors, but also in overinclusion errors that are attributable to sorting by a broader concept than the instructed one.

A similar question is whether the overexclusion errors resulted from the use of excessively narrow concepts. If this were the case, any subject who excludes would be expected to sort by only one of the subconcepts that the experimenter built into the concept specified for sorting. However, there is no basis on which to predict which of the several such subconcepts the subject would use for sorting; and there is no conventional statistical technique to test the relevant contention. A qualitative inspection of the data indicated that the prediction was probably not confirmed, since for most individual subjects there did not appear to be among the various subcategories any differential excluding. Therefore, no extended attempt was made to obtain a suitable statistical test.

Comparison of schizophrenic and brain damaged subjects. As the foregoing indicates, all of the effects which were predicted and found for the schizophrenics were also found for the most part among the brain damaged patients. However, it is of interest to compare the schizophrenic and brain damaged on their relative overall tendencies to overinclude or overexclude. In order to measure this, all of the available data were used, including the "additional tasks." Altogether, there were 240 cards which could be incorrectly included and 240

TABLE 1

MEAN TOTAL OVERINCLUSION AND OVEREXCLUSION ERROR SCORES ON EACH SET OF TASKS

Level	Overinclusion Errors			Overexclusion Errors		
	Overinclusion tasks	Overexclusion tasks	Total	Overinclusion tasks	Overexclusion tasks	Total
Schizophrenic						
I[a]	10.57	1.61	12.18	1.39	2.61	4.00
II	6.85**	2.08	8.93**	2.40*	8.06**	10.46**
Brain damaged						
I[a]	3.23	.15	3.38	1.49	1.08	2.57
II	.36*	.33	.69	1.00	5.08*	6.08*
Normal						
I[a]	2.12	.12	2.24	.61	1.52	2.13
II	.42	.33	.75	.94	2.94	3.88

[a] The Level I scores are doubled to make them comparable to Level II scores.

* The difference between levels is significant at the 5% level.

** The difference between levels is significant at the 1% level.

TABLE 2

MEAN OVERINCLUSION ERROR SCORES ON THE OVERINCLUSION TASK FOR ITEMS WHICH WERE SIMILAR AND DISSIMILAR TO THE INSTRUCTED CONCEPT

	Schizophrenic	Brain damaged	Normal
Level I[a]			
Similar	9.39	3.23	2.06
Dissimilar	1.18	0	.06
Level II			
Similar	5.56	.10	.33
Dissimilar	1.29	.26	.09

[a] The Level I scores are doubled to make them comparable to Level II scores.

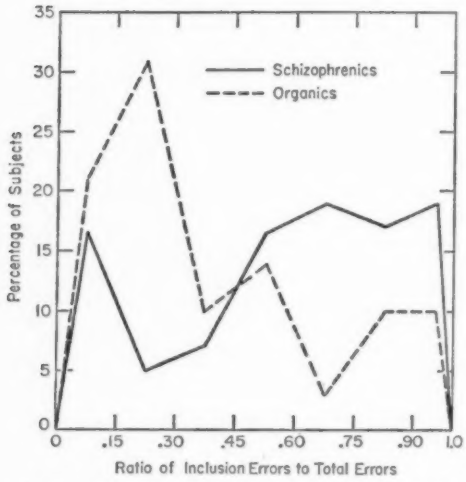


FIG. 1. Comparison of schizophrenic and brain damaged subjects on preference for overinclusion or overexclusion errors.

which could be incorrectly excluded. The score used was the proportion of total errors which were inclusion errors. Figure 1 shows the distributions of scores of schizophrenic and brain damaged patients on this index. In computing the data for Figure 1, those subjects were omitted who made fewer than five errors in all because a proportion score is highly unreliable for very low error scores. These omissions left a total of 29 brain damaged and 42 schizophrenic patients.

As indicated in Figure 1, 70% of the schizophrenics made more overinclusion errors than overexclusion errors, whereas 66% of the brain damaged showed the opposite preference, i.e., made more overexclusion errors than overinclusion. Note, however, the apparent bimodality of the schizophrenics' distribution. Although the schizophrenics, as a group, were significantly more overinclusive than the brain damaged ($p < .01$), 17% (seven of them) had a high proportion of errors on the overexclusion side.

An attempt was made to compare these seven patients on other variables with the main body of the schizophrenic group. They were not found to differ on age, education, chronicity, diagnostic subtype, or variety of symptoms. There was some suggestion that the seven overexcluders tended to be more effective hospital workers than the main body of schizophrenics.

DISCUSSION

The main hypothesis was confirmed in that the schizophrenics decreased their errors of overinclusion and increased their errors of overexclusion as the concepts specified by the instructions became broader. The same effects were also found for the most part with the brain damaged patients. Thus, for both kinds of patients there is a tendency to sort by concepts of a certain preferred breadth which accounts, in part, for both overinclusion errors and overexclusion errors. However, the tasks may differ from one another in ways other than conceptual breadth.

It is interesting to note, nevertheless, that the two groups of patients differed in relative tendencies toward overinclusion and overexclusion errors. This may reflect a difference between the groups on breadth of preferred sorting, or it may instead reflect tendencies toward overinclusion or overexclusion in addition to that arising from a preferred breadth. The first interpretation is consistent with McGaughran and Moran's (1957) finding that, when allowed to choose their own concepts on an object sorting test, brain damaged patients use narrower concepts than schizophrenics.

The normal subjects also showed differences between task levels of the same sort as the patients, which, although not significant, suggest the possibility that the effects observed for the pathological groups represent an increase of a normal error tendency. This is consistent with evidence pointed out by Chapman (1958) that many features of schizophrenic thinking disorder are accentuations of normal error patterns.

SUMMARY

The present study was designed to investigate the possibility that both overinclusion and overexclusion errors on conceptual sorting tasks by schizophrenics are attributable, at least in part, to a tendency to use concepts of a specific preferred breadth regardless of their appropriateness. This suggestion was investigated by means of a task in which names of objects were sorted on the basis of conceptual categories of varying breadth. The hypothesis

was that as the instructed conceptual category becomes broader, schizophrenics increase their errors of overexclusion but decrease their errors of overinclusion.

The hypothesis was confirmed. In addition, a group of brain damaged patients to a large extent showed the same effect, whereas a normal group showed only qualitative nonsignificant trends in this direction. However, the schizophrenic group on all tasks combined made more overinclusion errors than overexclusion, and the brain damaged showed the opposite error preference.

REFERENCES

CAMERON, N. Schizophrenic thinking in a problem-solving situation. *J. ment. Sci.*, 1939, **85**, 1012–1035.

CHAPMAN, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 374–379.

CHAPMAN, L. J., & TAYLOR, JANET A. The breadth of deviate concepts used by schizophrenics. *J. abnorm. soc. Psychol.*, 1957, **54**, 118–123.

EPSTEIN, S. Overinclusive thinking in a schizophrenic and a control group. *J. consult. Psychol.*, 1953, **17**, 384–388.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

KUGELMASS, S., & FOUNDEUR, M. P. Zaslow's test of concept formation: Reliability and validity. *J. consult. Psychol.*, 1955, **19**, 227–229.

MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44–49.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, 1944.

ZASLOW, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *J. consult. Psychol.*, 1950, **14**, 335–339.

(Received April 12, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 520-527

THE GENERALITY OF THE WAIS STANDARDIZATION FOR THE AGED: A REGIONAL COMPARISON[1]

CARL EISDORFER AND LOUIS D. COHEN
Duke University and Duke Medical Center

In establishing age norms for the Wechsler Adult Intelligence Scale (WAIS), some "1700 subjects, both sexes, aged 16–64 plus an additional 475 subjects of both sexes, ages 60–75 and over" were used. In selecting the 1700 subjects, "occupation and education, as well as sex and age, . . . also geography, urban *vs.* rural and color" were simultaneously considered, and it was concluded that "the norms as they stand . . . seem to be reasonably representative of the country as a whole . . . " (Wechsler, 1958, pp. 86–88, 91). The additional 475 subjects aged 60–75 and over referred to here formed the basis for an extension of the WAIS norms beyond age 64, but caution was expressed about using tables derived from these data because they were "not based on cases tested throughout the country according to census specifications but came from a special old age study" (Wechsler, 1955, p. 25). However, there was an overlap of groups aged 60–64 in both the national standardization and in the special old-age study, which was conducted in Kansas City. The overlapping groups showed high comparability of performance with only one subtest, Picture Completion, indicating a reliable difference between groups at the .05 level (Doppelt & Wallace, 1955).

The generality of the Kansas City results for the United States' population over 65 has needed testing, and the opportunity to make comparisons with a population in another part of the country arose in the course of an investigation conducted at Duke University in Durham, North Carolina, in connection with a longitudinal study of the aged. In a previous paper (Eisdorfer, Busse, & Cohen, 1959) some question was raised as to the generality of the Kansas City results in view of strikingly consistent discrepancies between the Verbal and Performance IQs in the Duke Project population when the Kansas City data were used for comparison. Using these latter criteria, it would have been expected that the relationship between Verbal and Performance IQs in the Duke data would be random, neither one nor the other being consistently superior; but successive comparisons based on age, sex, race, socioeconomic status, high or low IQ status, and community vs. hospitalized status in the Duke Project data showed consistent and reliable differences favoring the Verbal over the Performance IQs.

In consequence, it seemed advisable to compare systematically the data obtained from the final population of subjects available in the Duke study with those from Kansas City. The present report is based upon this comparison.

METHOD

Sample

The Kansas City study set out to test 500 individuals, although some 659 were listed to be approached on the basis of a "two stage, stratified, systematic sample" in which every block in the area of four counties embracing Kansas City was geographically numbered and weighted by size, and a systematic primary sample of every *n*th block selected. Every *n*th dwelling in each *n*th block was selected further, and each person over 60 in these dwellings was selected for the study. Actually, the WAIS was administered to 475 subjects, of whom 352 completed the study. Only data from completed tests were used for the IQ tables—a total of 352 subjects (Wechsler, 1955, p. 23).

The Duke sample of 239 subjects who completed the full scale WAIS was drawn from volunteers in the community, originally recruited through requests on radio and television as well as through Golden Age Club meetings. The census characteristics of Durham for age, sex, race, and occupation served as criteria for the recruitment of additional subjects to represent the various segments of the population in Durham city and county. The sample was built up over a 4-year period.

Actually 336 subjects were examined. Two institutionalized groups in Durham (15 in a retirement home and 64 in a state hospital) are not included in the present analysis. There were 18 subjects who did not complete the full scale WAIS. These too are not included. Our report is therefore based on 239 subjects.

Certain biases entered into the selection of the Duke

[1] This investigation was supported in part by Research Grants M900C, H3582, M2109 of the National Institutes of Health, United States Public Health Service. The help of E. W. Busse is gratefully acknowledged.

TABLE 1

COMPARISONS OF NUMBERS OF SUBJECTS AND AGE, RACE, AND SEX CHARACTERISTICS OF THE DUKE PROJECT AND KANSAS CITY STUDIES ON THE WAIS

Age	Duke Project							Kansas City Study[a] Total Sample		
	White		Negro		Total Sample					
	Male	Female	Male	Female	Male	Female	Total	Male	Female	Total
60–64	15	16	5	11	20	27	47	44	57	101
65–69	19	26	11	11	30	37	67	42	44	86
70–74	25	21	7	9	32	30	62	38	42	80
75+	23	20	9	11	32	31	63	36	49	85
Subtotal	82	83	32	42	114	125	239	160	192	352
Total	165		74		239					

[a] Wechsler (1955, Table 15, p. 23).

Project sample. Since all subjects had to come to the Medical Center for a 2-day examination, only ambulatory, reasonably well functioning individuals participated in the study. Bedridden, immobile, and socially isolated individuals are probably not represented in relation to their incidence in the Durham population. Since the sample was built up over time, many volunteers were not included in the study because they did not fit into the "cells" of the design that needed to be filled. Thus, toward the latter part of the study, "upper" socioeconomic Negro male volunteers were sought for and accepted, while "lower" socioeconomic white female volunteers were not. It may well be that some systematic bias was incurred by the inclusion of subjects who volunteered early, since those individuals may have been more socially alert and more assertive than others of their age in the Durham population.

In Table 1 may be seen the distribution of the subjects included in the Duke Study according to age, sex, and race characteristics. This distribution is compared with the Kansas City population on the same criteria. The total number of subjects in the Duke sample for the three age intervals 65–69, 70–74, and 75 plus is 67, 62, 63; for the Kansas City sample, $N = 86$, 80, and 85. The Duke and Kansas City samples for the three older groups seem close enough to each other in number to warrant comparison. Both studies report a slightly higher number of female subjects in the sample. Although the Durham population characteristics show 57% females over age 60, our Duke sample shows 52% female. The white/Negro proportions are not mentioned in the Kansas City study, but in the Duke sample, 31% are Negro, which conforms to 29% Negro in the Durham population. Some of the sample characteristics of the Duke Project are attributable to placing sufficient numbers of subjects in each cell so that individual variations would not distort the values of any cell.

No data on social class are reported for the Kansas City sample, but subjects were dichotomized on this variable in the Duke Project on the basis of a modification of Warner's Index of Status Characteristics (Warner, Meeker, & Eells, 1957) described in detail earlier (Eisdorfer et al., 1959). The "higher" status group included 103 subjects, whereas the "lower" socioeconomic group included 149 subjects.

Procedure

The WAIS examinations were administered by one of three experienced examiners according to the standard procedure outlined in the WAIS manual (Wechsler, 1955). The results for each subject were treated as prescribed in the test manual, Scaled scores, Verbal, Performance, and Full Scale IQs were computed.

In order to assess the relationship between the performance of the Duke and the Kansas City samples, comparisons were made by utilizing the data on Kansas City subjects reported by Doppelt and Wallace (1955).

RESULTS

Table 2 presents the Means and *SD*s of the Scaled scores for each of the WAIS subtests as well as for Verbal, Performance, and Full Scale scores and IQs. To assess the relationship between these scores and the data reported for Kansas City, *t* tests were performed to compare the Full Scale, Verbal, and Performance scores reported by Doppelt and Wallace (1955), and the data obtained in the present study. Table 3 gives the results of this comparison.

There is evidently a larger variance in the Duke sample, but no reliable difference between the Full Scale scores was obtained from the two groups. The import of the slight differences between the Full Scale scores of the two groups becomes clear on study of Sections B and C of Table 3. In Section B, subjects in age groups 70–74 and 75 plus in the Duke sample show significantly higher Verbal scores than do the subjects in the Kansas City sample. In Section C, we note that the Performance score of the 65–69 year old group of the Duke sample is significantly lower than the corresponding score in the Kansas City sample. It is interesting to observe also that in age group 60–64,

the Duke sample has a somewhat higher Verbal score and a somewhat lower Performance score. All of these comparisons confirm the previous report of Eisdorfer et al. (1959) of the consistent superiority of Verbal over Performance scores in an earlier Duke sample as compared with the Kansas City sample.

A more intensive analysis of the Verbal scale subtest scores is possible since Doppelt and Wallace (1955) present the Means and *SD*s of all their Verbal scale subtests. A comparison of the Verbal subtest performance of the Duke sample with that of the Kansas City sample is found in Table 4, which presents these data as well as the *t* tests for differences between the two samples.

Of special note is the Comprehension score of the Duke sample, which is significantly higher ($t < .01$) at every age level. Arithmetic and Vocabulary subtest scales also appear

TABLE 2

MEANS AND STANDARD DEVIATIONS OF ALL WAIS SUBTEST SCALED SCORES AND FULL SCALE, VERBAL, AND PERFORMANCE IQS FOR DUKE PROJECT SAMPLES

	Age Group 60–64 ($N = 47$)		Age Group 65–69 ($N = 67$)		Age Group 70–74 ($N = 62$)		Age Group 75 + ($N = 63$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Information	9.76	4.02	9.19	3.46	10.16	3.82	9.12	4.65
Comprehension	11.63	4.71	10.74	3.62	10.75	3.99	9.74	4.06
Arithmetic	9.76	4.31	8.67	3.33	9.08	3.45	8.28	3.49
Similarities	8.72	4.32	7.17	3.75	7.62	4.15	6.65	4.02
Digit Span	9.34	3.40	7.91	2.73	8.11	2.76	7.41	2.82
Vocabulary	10.65	4.24	9.85	3.89	10.56	4.38	9.90	4.70
Digit Symbol	5.14	3.54	4.43	2.92	4.37	3.22	2.73	2.66
Picture Completion	6.97	2.60	6.05	3.11	5.91	3.16	5.52	3.36
Block Design	6.42	3.20	5.71	2.61	6.04	2.89	5.14	3.23
Picture Arrangement	5.93	2.58	5.31	2.46	5.25	2.70	4.38	2.84
Object Assembly	6.97	3.18	6.05	2.56	6.16	2.75	5.20	2.73
Verbal score	59.89	22.78	53.70	18.12	56.14	20.00	50.96	21.27
Performance score	31.46	12.95	27.61	11.62	27.75	12.43	23.00	13.23
Full Scale score	91.23	34.68	81.31	28.89	83.90	31.24	73.74	33.36
Verbal IQ	104.00	22.56	100.01	17.95	108.20	19.99	106.77	21.01
Performance IQ	92.68	16.92	91.26	15.51	97.72	16.58	97.69	17.17
Full Scale IQ	99.08	20.68	96.04	17.35	104.00	18.86	102.88	19.79

TABLE 3

MEANS, STANDARD DEVIATIONS, AND *t* TESTS OF FULL SCALE, VERBAL, AND PERFORMANCE SCORES BETWEEN DUKE PROJECT AND KANSAS CITY SAMPLES[a]

	Duke Project Sample			Kansas City Sample			*t*	*p*
	N	*M*	*SD*	*N*	*M*	*SD*		
A. Full Scale score								
60–64	47	91.23	34.66	101	90.21	24.15	.18	*ns*
65–69	67	81.31	28.89	86	88.13	22.92	1.57	*ns*
70–74	62	83.90	31.27	80	77.19	21.42	1.43	*ns*
75+	63	73.74	33.36	85	68.71	21.56	1.03	*ns*
B. Verbal score								
60–64	47	59.89	22.78	101	55.24	14.51	1.27	*ns*
65–69	67	53.70	18.12	86	53.73	14.51	.01	*ns*
70–74	62	56.14	20.00	80	47.66	13.73	2.83	.05
75+	63	50.96	21.27	85	42.02	14.16	2.23	.05
C. Performance score								
60–64	47	31.46	12.95	101	34.97	10.94	1.59	*ns*
65–69	67	27.61	11.62	86	34.40	10.05	3.77	.01
70–74	62	27.75	12.43	80	29.53	9.45	.92	*ns*
75+	63	23.00	13.23	85	24.68	9.54	.84	*ns*

[a] As reported by Doppelt and Wallace (1955, Table 15, p. 328).

TABLE 4

MEANS, STANDARD DEVIATIONS, AND *t* TESTS OF DUKE PROJECT AND KANSAS CITY SAMPLES ON WAIS VERBAL SUBTESTS[a]

	Duke Sample			Kansas City Sample			*t*
	N	*M*	*SD*	*N*	*M*	*SD*	
Information							
60–64	47	9.76	4.02	101	9.55	3.17	.320
65–69	67	9.19	3.46	86	9.63	3.22	.791
70–74	62	10.16	3.82	80	8.48	2.86	2.867**
75+	63	9.12	4.65	85	7.92	2.81	1.811
Comprehension							
60–64	47	11.63	4.71	101	8.96	2.27	3.661**
65–69	67	10.74	3.62	86	8.76	2.43	3.830**
70–74	62	10.75	3.99	80	7.98	2.09	4.930**
75+	63	9.74	4.06	85	7.52	2.46	3.823**
Arithmetic							
60–64	47	9.76	4.31	101	9.51	3.58	.350
65–69	67	8.67	3.33	86	9.30	3.69	1.096
70–74	62	9.08	3.45	80	7.86	3.28	2.118*
75+	63	8.28	3.49	85	7.11	2.60	2.232*
Similarities							
60–64	47	8.72	4.32	101	8.77	2.82	.066
65–69	67	7.17	3.75	86	8.45	3.00	2.247
70–74	62	7.62	4.15	80	7.40	2.75	.371
75+	63	6.65	4.02	85	6.19	3.33	.735
Digit Span							
60–64	47	9.34	3.40	101	8.82	2.89	.898
65–69	67	7.91	2.73	86	8.01	2.69	.223
70–74	62	8.11	2.76	80	7.70	2.82	.868
75+	63	7.41	2.82	85	6.94	3.24	.939
Vocabulary							
60–64	47	10.65	4.24	101	9.61	3.25	1.487
65–69	67	9.85	3.89	86	9.58	2.83	.475
70–74	62	10.56	4.38	80	8.25	2.93	3.554**
75+	63	9.90	4.70	85	8.35	3.11	2.263*

[a] Doppelt and Wallace (1955, Table 9, p. 321).
* $p < .05$.
** $p < .01$.

higher in the Duke sample, but the differences are significant only for the two oldest groups—70–74, and 75 plus.

In 19 of the 24 comparisons of the Verbal subtest scores, the Duke sample is higher, 9 of the 19 being significant at the .05 level or better. The two older groups in the Duke sample are higher in every comparison. In only one case, the Similarities score of the 65–69 year old group, does the comparison go in the opposite direction at a significant level.

While Doppelt and Wallace (1955) presented the Means and *SD*s of all Verbal scale subtests for all subjects according to age, their presentation of the Performance scale subtests was not organized in the same manner. They list Performance scale data for the four age groups separately for men and women. Thus, it was necessary to treat data from the present study similarly. Table 5 contains the Performance scale scores of the men in the Duke and Kansas City samples, and Table 6 shows the mean Performance scale scores of women from the two samples.

The most striking finding appears in Table 5 in the age group 65–69, where the males in the Kansas City sample have higher scores than the corresponding Duke sample in four of the five performance subtests at a significance level of $p < .05$. The Kansas City females in the age group 65–69 are significantly higher on two of the subtests, Block Design and Object Assembly (in both cases $p < .01$). Tables 5 and 6 reveal a consistently poorer Performance score on the part of the Duke sample; of the 40 comparisons in Tables 5 and 6, 31 favor the Kansas City sample. The differences are not large except for the 65–69 year group.

In the treatment of their data, Doppelt and Wallace (1955, Table 7, p. 319) present different *N*s on the different Performance subtests for each age group. No explanation of this variation in *N*s is offered in the text, but one may assume that all the subjects completing the subtest were used, even though some did not complete the entire WAIS, but WAIS norms for these age groups are reported to have been computed from data obtained from subjects completing all the subtests (Wechsler, 1958). Thus, the Performance subtest data offered by Doppelt and Wallace (Table 7, p. 319) do not correspond to the data incorporated in the WAIS standardization, nor does the criterion information appear to be presently available in published form. It seems clear, therefore, that the Performance subtest comparisons between the Duke and Kansas City samples are based upon data closely approximating the criterion data (collected from Kansas City) and used as the WAIS norms for the aged.

Correlation matrices of the tests were prepared and are shown in Tables 7–10. These findings are in substantial agreement with those reported by Doppelt and Wallace (1955). As they reported, intercorrelations are generally higher among Verbal subtests and there appears to be similarity among the age groups as the correlation pattern emerges. Thus, these findings tend to support the contention of Dop-

TABLE 5

DIFFERENCES BETWEEN MEAN SCALED SCORE OF MEN IN THE DUKE AND KANSAS CITY SAMPLES ON EACH PERFORMANCE TEST[a]

Test	Age	Duke Sample			Kansas City Sample			*t* tests of difference
		N	*M*	*SD*	*N*	*M*	*SD*	
Digit Symbol	60–64	20	5.90	3.09	50	4.68	2.80	1.49
	65–69	30	3.56	2.66	44	5.07	1.83	2.65*
	70–74	32	4.31	3.03	44	3.00	2.50	1.97
	75+	32	3.03	2.55	40	2.40	2.40	1.05
Picture Completion	60–64	20	8.05	2.45	50	7.34	2.48	1.06
	65–69	30	6.16	3.14	47	7.47	2.48	1.89
	70–74	32	6.31	2.91	47	6.87	2.65	.85
	75+	32	6.18	3.59	50	5.32	2.81	1.14
Block Design	60–64	20	6.30	3.51	46	7.52	3.13	1.31
	65–69	30	4.80	2.59	43	7.42	2.39	4.32**
	70–74	32	5.90	2.96	44	6.02	2.62	.17
	75+	32	4.93	3.30	42	4.69	2.41	.35
Picture Arrangement	60–64	20	6.00	2.95	45	7.29	2.46	1.67
	65–69	30	5.00	2.44	43	6.86	2.51	3.11**
	70–74	32	5.34	2.78	41	6.37	2.51	1.60
	75+	32	4.59	3.09	40	5.05	2.37	.67
Object Assembly	60–64	20	6.65	3.01	46	7.33	2.77	.84
	65–69	30	5.73	2.30	43	7.98	2.76	3.72**
	70–74	32	5.87	2.56	40	6.25	2.95	.56
	75+	32	5.50	2.67	41	5.12	2.66	.59

[a] Doppelt and Wallace (1955, Table 7, p. 319).
* Significant at .05.
** Significant at .01.

TABLE 6

DIFFERENCES BETWEEN MEAN SCALE SCORES OF FEMALES IN THE DUKE AND KANSAS CITY SAMPLES ON EACH PERFORMANCE TEST[a]

Test	Age	Duke Sample			Kansas City Sample			*t* tests of difference
		N	*M*	*SD*	*N*	*M*	*SD*	
Digit Symbol	60–64	27	4.64	3.74	60	5.72	2.83	1.33
	65–69	37	5.13	2.98	49	4.53	2.59	.97
	70–74	30	4.37	3.50	46	3.41	2.02	1.33
	75+	31	2.41	2.77	61	3.02	2.29	1.02
Picture Completion	60–64	27	6.17	2.40	60	6.93	2.17	1.38
	65–69	37	5.97	3.13	54	6.54	2.49	.90
	70–74	30	5.48	3.47	51	5.35	2.48	.17
	75+	31	4.83	3.01	63	5.25	2.07	.67
Block Design	60–64	27	6.50	2.96	59	7.83	2.69	1.98
	65–69	37	6.45	2.42	49	7.61	2.69	2.05*
	70–74	30	6.20	2.90	49	6.55	2.53	.52
	75+	31	5.35	3.20	59	5.54	2.81	.26
Picture Arrangement	60–64	27	5.89	2.28	59	6.64	2.54	1.35
	65–69	37	5.56	2.47	46	6.30	2.27	1.37
	70–74	30	5.13	2.69	48	5.48	2.09	.57
	75+	31	4.16	2.59	50	4.98	2.32	1.41
Object Assembly	60–64	27	7.28	3.29	57	7.96	3.13	.88
	65–69	37	6.32	2.76	46	8.00	2.91	2.64*
	70–74	30	6.37	2.98	45	7.24	2.85	1.21
	75+	31	4.90	2.79	55	5.51	2.93	.93

[a] Doppelt and Wallace (1955, Table 7, p. 319).
* Significant at .05.

TABLE 7
INTERCORRELATIONS OF THE TESTS: AGE 60–64 DUKE SAMPLE[a]
($N = 47$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Information														
2. Comprehension	.819													
3. Arithmetic	.830	.855												
4. Similarities	.814	.800	.803											
5. Digit Span	.566	.681	.683	.669										
6. Vocabulary	.875	.877	.850	.884	.723									
7. Digit Symbol	.819	.749	.823	.807	.547	.805								
8. Picture Completion	.717	.769	.739	.682	.457	.717	.746							
9. Block Design	.685	.721	.655	.750	.508	.698	.655	.620						
10. Picture Arrangement	.623	.671	.679	.654	.502	.655	.653	.662	.613					
11. Object Assembly	.678	.704	.585	.633	.492	.645	.652	.571	.728	.731				
12. Verbal score[b]	.906	.932	.927	.917	.782	.961	.842	.758	.744	.699	.692			
13. Performance score[b]	.829	.845	.816	.830	.589	.827	.877	.831	.853	.843	.866	.876		
14. Full Scale score[b]	.903	.928	.913	.912	.733	.939	.878	.806	.805	.775	.779	.983	.947	
15. Verbal IQ	.906	.926	.923	.917	.783	.964	.840	.757	.742	.700	.695			
16. Performance IQ	.821	.836	.809	.830	.588	.830	.875	.822	.847	.838	.871	.872		
17. Full Scale IQ	.902	.922	.908	.910	.732	.941	.879	.802	.803	.770	.784	.981	.947	

[a] Compare with Doppelt and Wallace (1955, Table 10, p. 324).
[b] Correlation of tests with Verbal, Performance, and Full Scale scores before correction for contamination.

TABLE 8
INTERCORRELATIONS OF THE TESTS: AGE 65–69 DUKE SAMPLE[a]
($N = 67$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Information														
2. Comprehension	.846													
3. Arithmetic	.767	.679												
4. Similarities	.769	.818	.686											
5. Digit Span	.503	.576	.493	.396										
6. Vocabulary	.860	.837	.700	.725	.551									
7. Digit Symbol	.767	.703	.632	.686	.496	.774								
8. Picture Completion	.787	.769	.719	.706	.511	.729	.719							
9. Block Design	.625	.632	.569	.603	.423	.639	.656	.589						
10. Picture Arrangement	.719	.685	.634	.733	.370	.729	.652	.643	.627					
11. Object Assembly	.572	.586	.557	.628	.324	.548	.594	.707	.674	.533				
12. Verbal score[b]	.929	.927	.839	.860	.649	.914	.791	.827	.685	.765	.636			
13. Performance score[b]	.828	.802	.740	.794	.509	.813	.866	.876	.834	.812	.826	.881		
14. Full Scale score[b]	.916	.904	.824	.859	.612	.901	.845	.871	.766	.807	.731	.981	.955	
15. Verbal IQ	.927	.924	.840	.867	.639	.908	.792	.828	.681	.768	.641			
16. Performance IQ	.818	.792	.737	.799	.491	.803	.862	.868	.825	.813	.822	.873		
17. Full Scale IQ	.910	.897	.823	.866	.599	.894	.847	.867	.761	.809	.730	.976	.953	

[a] Compare with Doppelt and Wallace (1955, Table 11, p. 325).
[b] Correlation of tests with Verbal, Performance, and Full Scale scores before correction for contamination.

pelt and Wallace concerning the factorial composition of the WAIS for these age groups.

DISCUSSION

The differences which appear between the Duke Project sample and the standardization sample derived from Kansas City favor the Duke sample on the Verbal scale and, to a lesser extent, the Kansas City sample on Performance scale items.

The Duke 65–69 age group appears to differ most from the pattern of the Duke sample. As Table 3 shows, on the Full Scale, Verbal, and Performance comparisons, this group scored lower than the next older group in every case, when one might expect that it would perform at a midpoint between the 60–64 and 70–74 age groups. It can be seen from Table 3 that this expectation holds for the Kansas City sample. With the one exception of this Duke age group, all the age groups in both samples show decreasing levels of performance with increasing

TABLE 9
INTERCORRELATIONS OF THE TESTS: AGE 70–74 DUKE SAMPLE[a]
($N = 62$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Information														
2. Comprehension	.854													
3. Arithmetic	.793	.738												
4. Similarities	.762	.794	.573											
5. Digit Span	.577	.622	.584	.620										
6. Vocabulary	.876	.832	.708	.817	.630									
7. Digit Symbol	.772	.760	.670	.670	.563	.699								
8. Picture Completion	.819	.759	.688	.666	.576	.720	.731							
9. Block Design	.681	.640	.596	.605	.581	.593	.733	.736						
10. Picture Arrangement	.631	.642	.605	.499	.519	.706	.630	.604	.501					
11. Object Assembly	.569	.549	.403	.530	.475	.482	.577	.686	.699	.431				
12. Verbal score[b]	.926	.921	.831	.876	.746	.934	.796	.802	.705	.681	.576			
13. Performance score[b]	.830	.800	.708	.709	.646	.763	.880	.898	.873	.746	.801	.850		
14. Full Scale score[b]	.923	.907	.813	.842	.734	.901	.859	.870	.798	.732	.687	.977	.941	
15. Verbal IQ	.926	.914	.826	.875	.741	.939	.788	.809	.708	.683	.579			
16. Performance IQ	.827	.789	.705	.704	.640	.770	.864	.902	.867	.751	.796	.846		
17. Full Scale IQ	.921	.897	.809	.839	.730	.906	.846	.877	.798	.736	.685	.973	.940	

[a] Compare with Doppelt and Wallace (1955, Table 12, p. 326).
[b] Correlation of tests with Verbal, Performance, and Full Scale scores before correction for contamination.

TABLE 10
INTERCORRELATIONS OF THE TESTS: AGE 75+ DUKE SAMPLE[a]
($N = 63$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Information														
2. Comprehension	.874													
3. Arithmetic	.775	.754												
4. Similarities	.801	.749	.768											
5. Digit Span	.572	.559	.568	.541										
6. Vocabulary	.849	.857	.757	.837	.655									
7. Digit Symbol	.743	.712	.782	.745	.637	.780								
8. Picture Completion	.807	.767	.770	.798	.597	.795	.735							
9. Block Design	.625	.544	.625	.639	.525	.611	.716	.610						
10. Picture Arrangement	.773	.685	.690	.772	.552	.764	.768	.796	.734					
11. Object Assembly	.662	.555	.629	.727	.529	.662	.654	.683	.806	.723				
12. Verbal score[b]	.912	.896	.869	.900	.698	.946	.812	.861	.666	.803	.716			
13. Performance score[b]	.823	.748	.796	.838	.663	.824	.873	.875	.874	.904	.870	.883		
14. Full Scale score[b]	.904	.860	.865	.905	.696	.924	.855	.891	.770	.869	.803	.981	.953	
15. Verbal IQ	.910	.891	.869	.903	.700	.946	.816	.860	.672	.802	.721			
16. Performance IQ	.821	.742	.786	.834	.639	.820	.870	.872	.880	.908	.873	.875		
17. Full Scale IQ	.905	.859	.861	.903	.693	.923	.852	.890	.775	.862	.803	.980	.954	

[a] Compare with Doppelt and Wallace (1955, Table 13, p. 327).
[b] Correlation of tests with Verbal, Performance, and Full Scale scores before correction for contamination.

age. This 65–69 group has the largest N and the lowest SDs of the Duke sample.

In subtest comparisons, as noted on Tables 4, 5, and 6, the 65–69 group is the most atypical of the Duke sample. It is significantly lower in one Verbal subtest in comparison with the Kansas City sample—the only Duke subgroup that was significantly lower on a Verbal subtest—and in four of the six Verbal subtest comparisons, it was also lower. On the Performance subtests, it was not only lower than the comparable Kansas City group, but it was the only subgroup significantly lower than the comparison sample.

If we assume that the Duke 65–69 group is closer to the national standard and that the results for our other age groups are atypical, we are unable to explain the even lower Performance score of these subjects. At the very least, we must conclude that our subgroups are different from the Kansas City sample.

Since the criteria for the selection of groups

were similar for all ages, we cannot account for the performance of this group in terms of any systematic sampling bias. Since it included recently retired subjects, it perhaps involved a number of individuals who were under special emotional stress as a function of adaptation to their new status. This does not appear to have influenced the Kansas City sample, however, unless there may be differences in retirement age practices in the different regions.

It should be observed that this 65–69 group is not the principal contributor to the differences between the two samples. In fact, with regard to the Verbal differences noted, this group goes contrary to the Duke sample's higher achievement.

It is, of course, possible that the Duke Project sample is a biased one since all subjects were volunteers. On the other hand, it seems quite clear that this is also true of the Kansas City sample. Of the group originally contacted in Kansas City, approximately one-half (54.2%) completed the examination. Of the 649 subjects identified as the "Sample for testing," only 352 subjects completed the examination and were used for establishing the tentative national norms. In the Eisdorfer et al. (1959) study, using a portion of the Duke sample, various population variables were investigated in an attempt to find a source of sample bias. It appeared that the relationship between Verbal and Performance scores was not a function of sex, race, intelligence, social class, age, or mental illness variables. The authors can think of no reasonable explanation for the differences found, apart from unidentified regional differences.

If regional differences in patterns of intellectual performance exist, it would seem logical to suppose that they would derive from regional influences and would be intensified during the life span, thus appearing more prominently in the aged. This presumption could be tested readily. Probably in order to offset such (possible) regional influences, the collection of data from the WAIS norms for younger individuals (16–60) was based upon a national sample (Wechsler, 1958). In view of the present findings, it is suggested that subjects to be used as the standardization sample for the aged (60 plus) should also be selected on a nationwide basis.

SUMMARY

This study reports WAIS data on 239 elderly subjects above age 60 from the Durham, North Carolina, area. In view of the absence of national norms above age 65 for the WAIS and the suggested use of data from 352 subjects from a Kansas City study as tentative norms, it seemed especially appropriate to compare these latter with the Duke data. Stable differences were found between the samples, with the Duke groups showing superiority in the Verbal subtest scores and lower achievement in the Performance subtests. No satisfactory explanation of these differences is offered, other than the possibility of regional influences. The need for establishing national norms on a national sample of the aged is demonstrated.

REFERENCES

DOPPELT, J. E., & WALLACE, W. L. Standardization of the Wechsler Adult Intelligence Scale for older persons. *J. abnorm. soc. Psychol.*, 1955, **51**, 312–330.

EISDORFER, C., BUSSE, E. W., & COHEN, L. D. The WAIS performance of an aged sample: The relationship between verbal and performance IQs. *J. Geront.*, 1959, **14**, 197–201.

WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America.* Gloucester, Mass.: Peter Smith, 1957.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore: Williams & Wilkins, 1958.

(Received April 12, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 528-534

INTERPERSONAL PERCEPTIONS OF REPRESSORS AND SENSITIZERS AND COMPONENT ANALYSIS OF ASSUMED DISSIMILARITY SCORES[1]

JOHN ALTROCCHI
Duke University

One recent trend in relating individual differences in behavior to relatively enduring patterns of personality functioning has been the delineation of a dimension of repression-sensitization. Repressors, compared to sensitizers, have been shown, for instance, to have higher thresholds for material related to unacceptable needs (Eriksen, 1951) and to recall more completed than incompleted tasks under threat (Eriksen, 1954). Selection criteria for repressors and sensitizers have usually included the Hysteria (*Hy*) and Psychasthenia (*Pt*) scales of the MMPI and many investigators have related the repression-sensitization dimension to behavioral differences between patients using primarily repressive and denying defenses and those using primarily obsessive and intellectual defenses. From these studies the following definitions were derived (Altrocchi, Parsons, & Dickoff, 1960) and are used throughout the present study: Repressors are defined as those who tend to use avoidance, denial, and repression of potential threat and conflict as a primary mode of adaptation; sensitizers are defined as those who tend to be alerted to potential threat and conflict, to respond more readily with manifest anxiety, and to use intellectual and obsessive defenses.

In two experiments Gordon asked subjects to predict how their partners in the experiment would answer inventory questions. He found that sensitizers, as compared to repressors, predicted more dissimilarity between partners' responses and their own (Gordon, 1957). In a second study Gordon (1959) found that this difference was not modified by a 45-minute interaction or by the actual characteristics of the partner being judged. The present study is designed to clarify and extend Gordon's findings and their implications by seeking answers to the following questions:

1. Do sensitizers assume more dissimilarity than repressors between self and other in interpersonal perception (description of the other person, as distinguished from prediction of how the other person will answer items)?

2. If so, is the phenomenon evident in people who are better acquainted than Gordon's subjects, who were relative strangers?

3. Does the phenomenon extend to perceiving dissimilarity between two other people as well as to perceiving dissimilarity between oneself and others?

While these questions are meaningful in principle, Cronbach (1958) has recently criticized the use of assumed similarity (AS) or, more precisely, assumed dissimilarity (AD) scores which would usually be involved in seeking answers to them. He argues that "only after the simpler main effects associated with the perceiver or the object of perception have been given separate consideration" may AD scores be interpreted meaningfully as interactions between two sets of data (e.g., descriptions of self and other) (Cronbach, 1958). He suggests that AD scores can be analyzed into five simpler mathematical components—$\bar{X}$, σ_x, $\bar{Y}$, σ_y, and $r_{xy}\sigma_y/\sigma_x$—the means and standard deviations of the two descriptions and the regression of one on the other. Therefore a fourth question is asked:

4. Do differences in AD reflect a dyadic phenomenon, such as dependency of description of other on description of self, or can they be accounted for by differences in one or more of the monadic components (e.g., mean score on self-description)?

Another problem noted by Cronbach is that the usual form of the AD score combines two or more personality variables into a single

[1] This paper was presented at APA, September 1959. The study was carried out within the context of and was partially supported by Grant M-1599 from the National Institute of Mental Health. The author is indebted to the other investigators on the larger project and especially to Lee J. Cronbach, David M. Shaw, and Edward E. Jones for advice and criticism.

index which may obscure effects that involve some variables and not others. Therefore a final question is asked:

5. Are the differences in AD general or are they specific to certain descriptive personality dimensions?

METHOD

Subjects were selected from the total population of senior student nurses at the Duke University School of Nursing for 3 successive years (N = 227) on the basis of previously established MMPI criteria. These criteria —[$L + K + Dn$ (Denial subscale of Hy)] minus [$D + Pt$ (without K) + A]—were developed to increase reliability in selecting subjects on exactly the same dimension referred to by Gordon and previous investigators (Altrocchi et al., 1960). The top fourth of the population (N = 56, scores of +45 and above) were called repressors and the bottom fourth (N = 56, scores of −12 and below) were called sensitizers. Midgroup subjects (N = 57) were those with scores in the middle range who were above the median in ratings of adjustment (in order to have as normal a midgroup as possible). Subjects' adjustment had been rated from their MMPIs by experienced judges for another study.

All subjects described themselves and three randomly chosen students in their training group (Ns = 6-13) on the Interpersonal Check List (ICL) (Leary, 1957), which consists of 128 phrases to be checked true or false, 16 phrases falling into each of eight octants, each reflecting a different kind of interpersonal behavior. Subjects were tested before and after a 3–4 month period of training in psychiatric nursing. This intensive, interpersonally-oriented training included individual counseling aimed at helping the student become more aware of the dynamics underlying her own behavior in her relationships with her patients. Before the training the students were acquainted with each other to varying degrees but they had normally spent 3 years in school together. The pretraining data thus reflect a degree of acquaintance considerably beyond that of Gordon's laboratory subjects. During training the students interacted with the others they had to describe not only on the wards, but also in small seminars where realistic appraisals of self, patients, and peers were encouraged. The posttraining data, therefore, reflect not only further acquaintance, but possible effects of the training on interpersonal perception.

The AD scores were calculated in the usual manner for difference scores—i.e., by taking the square root of the sum (across octants of the ICL) of the squares of the differences between the two descriptions (self and other or two others) (Cronbach, 1958).

Two separate ex post facto designs were involved. The first was a mixed 2 × 2 × 2 design with N = 23 in each cell. There were pre- and post-AD scores for repressors and sensitizers involved in one of two kinds of tasks—describing someone similar to themselves (e.g., a repressor describing another repressor) or different from themselves. (The midgroup was omitted from this analysis because the squaring procedure involved in calculating AD scores gives greater weight to large differences, and midgroup subjects presumably have less chance to be greatly different from either the repressors or the sensitizers to be described.) The second design was a mixed 2 × 3 × 2 design with N = 12 in each cell. There were pre- and post-AD scores for three kinds of subjects (repressors, midgroup, and sensitizers) involved in one of two kinds of tasks. One task was to describe two people who were similar to each other—two repressors or two sensitizers. The other task was to describe two people who were different from each other—a repressor and a sensitizer.

TABLE 1
MEAN ASSUMED DISSIMILARITY (AD) SCORES FOR THE FIRST DESIGN

Group	Pretraining			Posttraining		
	Describing a person similar to oneself	Describing a person different from oneself	M	Similar	Different	M
Repressors	7.9	7.8	7.9	6.2	7.9	7.1
Sensitizers	11.5	11.0	11.3	9.6	10.3	10.0
Total	9.7	9.4	9.6	7.9	9.1	8.5

Note.—N = 23 in each cell.

TABLE 2
ANALYSIS OF VARIANCE OF AD SCORES FOR THE FIRST DESIGN

Source	SS	df	MS	F
Between subjects	2037.3	91		
Repressors-sensitizers (B)	455.2	1	455.2	25.6**
Person described (Similar-Different) (C)	8.8	1	8.8	<1
BC	5.0	1	5.0	<1
Error (b)	1568.3	88	17.8	
Within subjects	450.3	92		
Pre-Post (A)	54.8	1	54.8	13.0***
AB	2.9	1	2.9	<1
AC	25.6	1	25.6	6.1**
ABC	0.6	1	0.6	<1
Error (w)	366.4	88	4.2	
Total	2487.6	183		

** $p < .05$.
*** $p < .01$.

RESULTS

AD Analysis

The mean AD scores for the four groups pre and posttraining in the first design are

presented in Table 1. The results of a Lindquist (1953) Type III analysis of variance are presented in Table 2. The AD scores for sensitizers were significantly higher than the AD scores for repressors ($p < .01$). The AD scores were not significantly affected by whether the person described was actually similar or dissimilar to the subject. The significant interaction between pre-post and similar or different persons described ($p < .05$) suggests that subjects' perceptions more closely reflected actual differences and similarities after training than before. The AD scores also decreased significantly from pre to posttraining across all groups ($p < .01$).

Thus sensitizers do seem to assume more dissimilarity between self and other in interpersonal perception than repressors, even after extensive acquaintance and special clinical training.

Mean AD scores for all six groups in the second design (involving descriptions of two others) are presented in Table 3. Results of a Lindquist Type III analysis of variance are presented in Table 4. The only significant effect is that based on whether the two persons described actually were similar or dissimilar ($p < .05$). There was a trend, which did not reach an acceptable level of statistical significance ($p = .10$), for sensitizers to assume more dissimilarity between two others.

Component Analysis

To answer the fourth and fifth questions concerning component analysis, it was necessary to analyze the AD scores into the five components, ($\bar{X}$, σ_x, $\bar{Y}$, σ_y, and $r_{xy}\sigma_y/\sigma_x$),

TABLE 3
MEAN AD SCORES FOR THE SECOND DESIGN

Group	Pretraining			Posttraining		
	Describing two people similar to each other	Describing two people different from each other	*M*	Similar	Different	*M*
Repressors	6.9	9.2	8.1	6.6	9.0	7.8
Midgroup	8.8	8.3	8.5	7.2	7.4	7.3
Sensitizers	8.1	12.2	10.2	8.5	11.5	10.0
Total	7.9	9.9	8.9	7.4	9.5	8.4

Note.—$N = 12$ in each cell.

TABLE 4
ANALYSIS OF VARIANCE OF AD SCORES FOR THE SECOND DESIGN

Source	*SS*	*df*	*MS*	*F*
Between subjects	2541.5	71		
Repressors-sensitizers (B)	153.5	2	76.7	2.3*
Persons described (Similar-Different) (C)	139.8	1	139.8	4.3**
BC	85.1	2	42.5	1.3
Error (b)	2163.1	66	32.8	
Within subjects	539.6	72		
Pre-Post (A)	12.0	1	12.0	1.5
AB	7.9	2	3.9	<1
AC	0.0	1	0.0	<1
ABC	4.8	2	2.4	<1
Error (w)	514.9	66	7.8	
Total	3081.0	143		

* $p = .10$.
** $p < .05$.

noted by Cronbach (1958), for at least two independent dimensions of the ICL. Therefore scores were calculated for all 46 repressors and 46 sensitizers (from the first design) on the two moderately independent dimensions of the ICL—Dominance and Love (Leary, 1957)—for description of self and description of other separately. Data for description of others who were repressors and others who were sensitizers were combined. Dominance and Love were used rather than octants because the octants are intercorrelated and use of the octant scores would thus violate one of Cronbach's (1958) specifications for component analysis. The original AD scores, however, were calculated using octant differences. Thus it was necessary to demonstrate that AD scores based on Dominance and Love would show differences between repressors and sensitizers parallel to the differences found using AD scores based on octants. Therefore AD scores for self-other on Dominance pretraining were calculated. The scores for sensitizers were higher than those for repressors ($t = 3.12$, $p < .01$), suggesting that a component analysis of Dominance and, presumably, Love scores would be likely to reveal contributions to the differences in AD scores using octants.

Tables 5 and 6 present the results of the component analyses for pre and posttraining Dominance and Love scores. Before the training repressors described themselves with less variability and with higher scores than sensitizers did on Dominance and, to a lesser extent,

TABLE 5

COMPARISONS BETWEEN REPRESSORS AND SENSITIZERS ON DOMINANCE AND LOVE T SCORES FOR PRETRAINING DESCRIPTION OF SELF AND DESCRIPTION OF OTHER SEPARATELY

Component	Repressors	Sensitizers	F or t	p
$\bar{X}$ Dominance self	57.8	51.6	4.40[a]	<.001
σ^2 Dominance self	27.3	62.3	2.28	<.02
$\bar{X}$ Dominance other	60.8[b]	59.3[b]	<1	
σ^2 Dominance other	56.3	78.3	1.39	
r Self-other (Dominance)	.19	−.01	<1	
r Self-other (σ other/σ self)	.27	−.01		
$\bar{X}$ Love self	53.0	48.4	2.89[a]	<.01
σ^2 Love self	40.6	73.6	1.82	<.10
$\bar{X}$ Love other	51.2	48.6	1.05	
σ^2 Love other	137.0	139.0	1.01	
r Self-other (Love)	.23	.08	<1	
r Self-other (σ other/σ self)	.42	.11		

[a] To insure that these significant t tests did not result from the different variances for repressors and sensitizers, the distributions were split at the medians and chi square tests were significant at <.001 for Dominance and <.05 for Love.

[b] Difference between Dominance self and Dominance other for sensitizers $t = 4.5$, $p < .001$; for repressors $t = 2.4$, $p < .05$.

TABLE 6

COMPARISONS BETWEEN REPRESSORS AND SENSITIZERS ON DOMINANCE AND LOVE T SCORES FOR POSTTRAINING DESCRIPTION OF SELF AND DESCRIPTION OF OTHER SEPARATELY

Component	Repressors	Sensitizers	F or t	p
$\bar{X}$ Dominance self	58.6	53.3	3.80[a]	<.001
σ^2 Dominance self	28.7	57.9	2.02	<.05
$\bar{X}$ Dominance other	61.5	57.9[b]	1.81	
σ^2 Dominance other	56.8	115.6	2.04	<.05
r Self-other (Dominance)	−.20	.12	<1	
r Self-other (σ other/σ self)	−.28	.17		
$\bar{X}$ Love self	51.2	47.3	2.25[a]	<.05
σ^2 Love self	42.1	88.9	2.09	<.05
$\bar{X}$ Love other	49.0	49.3	<1	
σ^2 Love other	100.9	107.2	1.06	
r Self-other (Love)	.46***	.23	1.25	
r Self-other (σ other/σ self)	.71	.25		

[a] The distributions were split at the medians and a chi square test was significant at <.001 for Dominance but not for Love ($\chi^2 = 1.57$, $p < .20$).

[b] Difference between Dominance self and Dominance other for sensitizers $t = 2.5$, $p < .05$.

*** $p < .01$.

on Love. On the other hand, Table 5 discloses no evidence that repressors and sensitizers differed significantly in their perceptions of others. Furthermore, the low correlations and, therefore, low regression components in Table 5 suggest that neither in the repressor nor sensitizer groups did the description of others depend substantially on the description of self. Of the components of AD noted by Cronbach, regression of other description on self-description is the only dyadic component and comes closest to reflecting assumed similarity as ordinarily conceptualized. The differences between repressors and sensitizers in the correlations are in the direction consistent with the earlier AD analysis (in Tables 1 and 2) for Love before and after the training and for Dominance before the training. Thus there was a trend, although not statistically significant, for the repressors to assume more similarity than did sensitizers between self and others before the training.

After the training (see Table 6) approximately the same pattern was still apparent; repressors had higher and less variable scores in self-description, especially on Dominance. The slight decreases in the differences in self-description could reflect regressions to the mean or slight changes in self-concept resulting from the training. One comparison in descriptions of others did become significant after training; sensitizers were more variable ($p < .05$) in their descriptions of others on Dominance. Comparison between Tables 5 and 6 does not reveal any consistent changes in means or variances which would clearly explain the significant pre-post decrease in self-other AD scores (see Tables 1 and 2). One possible factor in this decrease might be the significant posttraining self-other correlation ($r = .46$, $p < .01$) for repressors on Love. Probably important in this decrease and in evaluating all pre-post changes, however, was a finding from analysis of these data for another study that subjects checked fewer items true in posttraining testing.

The component analyses give answers to Questions 4 and 5: The highly significant differences found in this study between repressors and sensitizers in self-other AD scores (see Tables 1 and 2) were due primarily to two of the monadic components—differences in means and standard deviations of their self-concepts. These differences were general enough to be consistent across two descriptive personality dimensions—Dominance and Love, as defined by Leary (1957). There was little evidence that either repressors or sensitizers tended to base their descriptions of others on their descriptions of themselves—and thus little evi-

dence of a dyadic phenomenon—and no clear evidence that the groups tended to differ significantly in their perception of others.

The component analyses raise questions as to whether the analysis of AD scores for description of two others (see Tables 3 and 4) actually answered the original question concerning perceptions of difference between others. It has become apparent that AD scores cannot be interpreted meaningfully unless they are analyzed into components. Yet comparing the mean and standard deviation components in this case would only indicate the adequacy of the selection of pairs of people to be described. Thus it seems that the closest one can come to comparing differences between descriptions of two others in the present design is to calculate the correlation component. The most uncontaminated measure of this component is covariance ($\Sigma xy/n$), which Bass and Fiedler (1959) used, but in this study the more familiar product-moment correlation coefficient ($\Sigma xy/n\sigma x\sigma y$) is used because it makes comparison between groups simpler. The correlations between Dominance scores for description of two others and Love scores for the same descriptions for all groups in the second design are presented in Table 7. Since only 2 of 24 correlations were significant, these results must be viewed cautiously. Some trends are apparent, though. The correlations tended to be higher in the positive direction for repressors than midgroup subjects, while the correlations for sensitizers tended to be negative. The correlations tended to be higher when describing two similar people than when describing two different people, higher before training, and higher for Love than for Dominance. Nevertheless, transforming all r's into z scores and comparing z's disclosed only two statistically significant comparisons; the correlation on Dominance, posttraining, describing two similar people for midgroup subjects was higher than that for repressors ($p < .05$) and sensitizers ($p < .01$). Thus these analyses did not provide clear support for the suggestion that sensitizers assume more dissimilarity between two others, but the trends in the results suggest some possible leads for future investigators.

TABLE 7
PRODUCT-MOMENT CORRELATIONS BETWEEN DESCRIPTIONS OF TWO OTHERS ON DOMINANCE AND LOVE

Variable and Group	Pretraining		Posttraining	
	Describing two people similar to each other	Describing two people different from each other	Similar	Different
Dominance				
Repressors	.32	.39	−.16	.03
Midgroup	.10	−.03	.79***	−.54
Sensitizers	−.28	.04	−.38	−.32
Love				
Repressors	.54	.30	.70***	.47
Midgroup	.50	.13	.37	.10
Sensitizers	.26	−.14	.13	−.41

Note.—$N = 12$ for each correlation.
*** $p < .01$.

DISCUSSION

The present results clearly support Cronbach's (1958) conclusion that AS and AD scores can be greatly misleading. Whereas the AD analyses might have led to interpretations regarding different interpersonal perception processes for repressors and sensitizers, the component analyses demonstrated no perceptual differences between the groups. Cronbach's conclusion receives further corroboration from a study by Bass and Fiedler (1959) in which AD scores did not consistently relate to their covariance components and self-concept differences were found to be responsible for some of Fiedler's earlier results using AS scores. Self-concept differences may also be responsible for the differences Gordon (1957, 1959) found between repressors and sensitizers in interpersonal prediction and for some of the results in other studies using assumed similarity scores (e.g., Rosenbaum, 1959). It is noteworthy that Bronfenbrenner (1958) has reached a similar conclusion regarding the use of gross similarity scores in the measurement of identification. The potential usefulness of Cronbach's component analysis as a substitute for use of AD scores seems demonstrated by the present study. Hopefully other methods such as multivariate r statistics and associated designs suggested by La Forge (1959) will also prove useful in studies of interpersonal perception.

Since AD scores can be misleading, Cronbach[2] concludes that statements of hypotheses in AD terms are valueless and tend to obscure

[2] Personal communication.

dependable relationships. This conclusion can be further supported by noting the logical difficulties in measuring a phenomenon such as assumed dissimilarity (or similarity). Such a phenomenon seems to imply measurement of the degree of agreement between two measures, for which the intraclass correlation coefficient (r_i) (Robinson, 1957) is a more appropriate statistic than r. Even r_i, however, would need to be analyzed into components involving means, standard deviations, and covariance to allow precise interpretation and only if covariance were the primary component would use of the term assumed similarity be appropriate. This investigator is led to conclude that development and use of concepts at least as precise as our measuring techniques will allow would greatly accelerate meaningful research in interpersonal perception.

The present results demonstrate, more clearly than had been shown earlier with many of the same subjects (Altrocchi et al., 1960), that repressors have more positive self-concepts than sensitizers (in this case describing themselves as higher on Dominance and Love) and that they are a more homogeneous group in self-description. These results are not surprising in that there is growing evidence that subjects with negative self-concepts show greater instability of self-concept (cf. Altrocchi, et al., 1960; Engel, 1959); and one would expect people who describe themselves negatively on one self-report test to do so on another. Many sensitizers evidently see themselves as different from and less desirable than other people (cf. Table 5) and do not simply describe everyone negatively, but it is not clear whether sensitizers tend to describe others differently from the way repressors describe others. Considering the heterogeneity of the sensitizer group, it might be useful for future investigators to define more homogenous sensitizer subgroups.

The only results in this study which are not likely to have been directly affected by self-concept differences between repressors and sensitizers are those involving perceived dissimilarity between two others (see Tables 3, 4, and 7). There was a slight tendency for sensitizers to perceive greater dissimilarity between two others as measured by correlations between descriptions of two others. It might be profitable to explore this difference further, since it seems possible that Fiedler's (1958) measure of "assumed similarity between opposites" (ASo) might be partly a function of such individual differences as those between repressors and sensitizers. Another reason for pursuing this problem is that Holzman and Gardner (1959) have recently demonstrated that extreme repressors tend to be levelers on a neutral psychophysical task. It would be useful to know whether we can link defensive processes, such as repression or sensitization, not only to adaptive processes in perception of physical stimuli, as Holzman and Gardner have done, but also to modes of interpersonal perception.

The present results are relevant for studies of changes in self-concept or perception of others as a function of special training. The clinical training of these nursing students did not result in increased differentiation (variability) in perception of peers of the kind that Crow (1957) found in medical students' perceptions of patients after training in accuracy of perception. The conditions under which clinical training results in changes in perceptions of self and others remain unclear. Finally, the present results suggested the need to take account, in future studies, of subjects' possible tendencies to check fewer items true on the ICL in a second testing.

SUMMARY

From a population of 227 student nurses, MMPI criteria were used to select 56 repressors and 56 sensitizers. Each subject described herself and three of her peers on the Interpersonal Check List before and after intensive training in psychiatric nursing. Assumed dissimilarity (AD) scores of independent groups of repressors and sensitizers describing other repressors and sensitizers were calculated. The results, following two studies by Gordon, seemed to demonstrate that sensitizers assume more dissimilarity between self and other than repressors. Following suggestions by Cronbach, however, AD scores were analyzed into components, revealing that the differences in assumed dissimilarity between self and others were due primarily to stable differences in self-description and not to any

clear differences in perception of others nor to any substantial correlation between perception of others and perception of self (the dyadic component). The results demonstrate the usefulness of Cronbach's suggestions and the necessity of analyzing gross AD scores into components in order to interpret them with precision. It is further concluded that use of such gross concepts as assumed dissimilarity tends to obscure relationships.

Sensitizers were again shown to have more negative self-concepts than repressors and to be a more heterogeneous group, suggesting the need to examine possible sensitizer subgroups. Other suggestions for further investigation stress the need to clarify, more than the present results did, whether sensitizers, more than repressors, tend to perceive dissimilarity between other people; whether there are, in fact, any differences between repressors and sensitizers in perception of others; and whether special clinical training does result in changes in perception of self or others.

REFERENCES

ALTROCCHI, J., PARSONS, O. A., & DICKOFF, HILDA. Changes in self-ideal discrepancy in repressors and sensitizers. *J. abnorm. soc. Psychol.*, 1960, **61**, 67–72.

BASS, A. R., & FIEDLER, F. E. Interpersonal perception scores: A comparison of D scores and their components. Technical Report No. 5, 1959, University of Illinois, Group Effectiveness Research Laboratory.

BRONFENBRENNER, U. The study of identification through interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Ch. 9.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Ch. 23.

CROW, W. F. The effect of training upon accuracy and variability in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **55**, 355–359.

ENGEL, MARY. The stability of the self-concept in adolescence. *J. abnorm. soc. Psychol.*, 1959, **58**, 211–215.

ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, **46**, 557–564.

ERIKSEN, C. W. Psychological defenses and ego strength in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

FIEDLER, F. E. Interpersonal perception and group effectiveness. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Ch. 16.

GORDON, J. E. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, 1957, **25**, 686–698.

GORDON, J. E. The stability of the assumed similarity response set in repressors and sensitizers. *J. Pers.*, 1959, **27**, 362–373.

HOLZMAN, P. S., & GARDNER, R. W. Leveling and repression. *J. abnorm. soc. Psychol.* 1959, **59**, 151–155.

LAFORGE, R. Rater variance in ratings of persons: The generality of the D^2-statistic. Paper presented at the meetings of the American Psychological Association, September 1959.

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

ROBINSON, W. S. The statistical measurement of agreement. *Amer. sociol. Rev.*, 1957, **22**, 17–25.

ROSENBAUM, M. E. Social perception and the motivational structure of interpersonal relations. *J. abnorm. soc. Psychol.* 1959, **59**, 130–133.

(Received April 19, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 535-542

DOMINANCE AS A FACTOR IN ACHIEVEMENT AND PERCEPTION IN COOPERATIVE PROBLEM SOLVING INTERACTIONS[1]

WILLIAM T. SMELSER

Institute of Human Development, University of California

The first part of this study relates different pairings of dominant and submissive males, interacting under various role assignments, with their joint achievement in a cooperative problem solving situation. The second part investigates the relationship of these different pairings to the perceived dominance of the self and the partner.

The thesis is advanced that different combinations of dominant and/or submissive individuals achieve more or less successfully according to the pair-combination as well as the conditions of assignment of dominant or submissive roles, and that it is possible to predict differential success among these permutations according to hypotheses derived from personality theory.

The hypotheses are derived from Sullivan's (1953) general assumption that a person's modes of relating to others are functional in that they enable him to maintain anxiety at a minimum. Interpersonal situations that permit, or encourage, the use of a salient interpersonal technique, such as dominance or submission, give rise to less anxiety than situations that do not. The experiencing of anxiety disrupts cognitive functioning and leads to a less effective performance in a task. Hence, there should be greater achievement when persons are permitted or encouraged to assume habitual modes of relating.

In spite of the explicitness and theoretical importance of this reciprocal aspect of interpersonal theory, few experimental studies have been concerned with it. Leary (1957), employing dominance as one of his central variables, posits a reciprocal theory of interaction. In Mann's (1959) review of the studies of the relation of personality and performance in small groups, he did not report any studies considering the relationship between interacting individuals differing in dominance and the influence of this relation upon either the nature of group achievement or mutual perception. Tuma (1955) found that dissimilarity in dominance between therapist and client, regardless of the direction of the difference, correlated highly with rated improvement in a counseling situation.

The following four hypotheses are based on a theoretical prediction of differences, among different pairings of male subjects under specified conditions of role assignment, in achievement in a cooperative task situation. Table 1 describes the seven different groupings and the interaction and the role assignments, when such assignments were made. The predicted rank orders of achievement of the groups involved in each of the four hypotheses are presented in Table 2.

Hypothesis 1 compares Groups A, B, and G. Following Sullivan (1953), the pairings in Group A perform the most productively, since each member of the pair has a salient interpersonal technique explicitly requested by the experimenter as well as "pulled for" by the behavior of the partner. The pairs of dominant and submissive subjects comprising Group B interact under somewhat less optimal conditions, since these pairs lack the congruent role assignments of the pairs in Group A. The pairs of dominant and submissive subjects in Group G interact with an assignment of roles contrary to their preferred modes, which arouses anxiety and disrupts achievement.

Hypothesis 2 compares Groups B, E, and F. Paired dominant subjects (Group E) do not interact under as optimal conditions as do the pairs in Group B, since each member of the paired dominant subjects interacts so as to induce submissive behavior in his partner, who is also a dominant subject. The paired subjects in Group E were predicted to perform more

[1] This paper is based on a dissertation submitted to the Department of Psychology, University of California, in partial fulfillment of the requirement of the PhD. The writer wishes to express his appreciation for their help to Read Tuddenham, Edwin Ghiselli, Harrison Gough, Frederick Tyler, and Lyman Porter.

successfully than the paired submissive subjects in Group F, since dominant subjects are hypothesized to behave in the pursuit of satisfaction rather than security, and thus to be more at ease with each other and themselves than the paired submissive subjects. Sullivan (1953) considered the power motive as one of the pursuits of satisfaction, rather than security. Individuals lacking in dominance may have found it necessary to dissociate the expression and/or awareness of this motive.

Hypothesis 3 compares Groups C, D, E, and F. The prediction involves the differential achievement of two different pairings (two dominant subjects interacting and two submissive subjects interacting), each under two conditions (no role assignments and assignment of the dominant role to one subject and assignment of the submissive role to the partner). The predicted rank order of achievement, by groups, is C, D, E, and F, since it was hypothesized that pairings with role assignment perform more successfully than comparable pairings with no role assignment, and further, that the paired dominant subjects would perform more successfully than paired submissive subjects under comparable conditions. For a dominant subject to relinquish a dominant role should produce less anxiety than for a submissive subject to assume a dominant role.

TABLE 1
PAIRINGS OF DOMINANT AND SUBMISSIVE SUBJECTS AND CONDITIONS OF ROLE ASSIGNMENTS
(10 pairs in each group)

Group	Description of Interaction	Code
A	Dominant subject assigned dominant role; submissive subject assigned submissive role.	D:d-S:s
B	Dominant subject and a submissive subject interact with no role assignment.	D:x-S:x
C	One dominant subject assigned dominant role; other dominant subject assigned submissive role.	D:d-D:s
D	One submissive subject assigned dominant role; other submissive subject assigned submissive role.	S:d-S:s
E	Two dominant subjects interact with no role assignment.	D:x-D:x
F	Two submissive subjects interact with no role assignment.	S:x-S:x
G	Dominant subject assigned submissive role; submissive subject assigned dominant role.	D:s-S:d

TABLE 2
HYPOTHESES OF RANK ORDER OF ACHIEVEMENT OF GROUPS

Hypothesis	Predicted Rank Order of Achievement
1	A > B > G
2	B > E > F
3	C > D > E > F
4	A > B > C > D > E > F > G

In Sullivan's (1953) terms, it is less disruptive to modify a mode of attaining satisfaction (dominant role) than it is to modify a security operation (submissive role).

Hypothesis 4 is not independent of the other three hypotheses, but compares achievement among all seven groups. The predicted rank order corresponds to the alphabetical ordering of the groups, i.e., Group A is predicted to be the most successful and Group G the least successful. Hypotheses 1 and 3 are independent, since their predictions do not involve any common groups. Hypothesis 2 has one group (Group B) in common with Hypothesis 1 and it has two groups in common with Hypothesis 3 (Groups E and F).

METHOD

The subjects were all male volunteers enrolled in either Military Science or Air Science at the University of California, Berkeley. Since enrollment in the university armed forces training program is compulsory, this group afforded a broad sampling of university undergraduate men. The total number of subjects who volunteered was 748. Participation in the experiment was credited as an excuse from one military drill period.

Before requesting volunteers, the students were administered a short form of Gough's California Psychological Inventory (CPI), which contained the Dominance (Do), Social Responsibility, and Self-Control scales (Gough, 1957). A vocabulary test (Thorndike & Gallup, 1944) was administered at the conclusion of the personality test. The subjects for the experiment were selected from the extremes of the distribution of the Do scores. The mean of this distribution was 28.5 (*SD*, 6.5) which is quite similar to Gough's (1957) standardization group mean of 28.5 (*SD*, 6.0). A subject scoring 34 or above was defined as dominant; a subject scoring 23 or below as submissive. The raw scores of 34 and 23 correspond to Gough's T scores of 110 and 92, respectively, on a scale with a mean of 100 and a standard deviation of 10.

Experimental Situation

The interaction of the paired subjects involved the

joint operation of two model railroad trains. The apparatus was that employed by Ghiselli and Lodahl (1958). The main track was 6 feet in diameter, and had two bypass sidings by which the two trains could leave and enter the main track. The two subjects sat on high stools to operate their respective control panels, which were 3 feet high, on the same level with the tracks. The left half of each control panel contained switches that supplied power to different numbered sections of the track. The right half of each control panel contained several controls: direction and speed of the trains, switches to and from the sidings, and a master control switch. The subjects were thus able to control the speed, direction, and route of each train. Carelessness by either subject in operation of his switches could obstruct the efforts of the partner in operating his train.

Procedure

Before the instructions for operation of the trains, the subjects were introduced, both as a matter of social convention, and also to determine whether the subjects were acquainted with each other. Subjects previously acquainted were not included in the results. Subjects from Military Science and Air Science were not mixed in the pairings. The instructions for operation of the two trains were given during the first 20 minutes. Both trains were operated by the experimenter at this time, and both control panels were equally employed for demonstration. The subjects were permitted to ask questions during the demonstration. They were allowed 3 minutes to discuss the operation of the trains before the first trial.

The role assignments, when given, were given just prior to the first trial. In the following example, "X" is assigned the dominant role and "Y" is assigned the submissive role.

Mr. X, your train will be the passenger train, and Mr. Y, your train will be the freight. Mr. X, you are the dispatcher and you will arrange and order the solutions to the problem, and Mr. Y, your task is to carry out, on your board, Mr. X's directions. Remember, Mr. X, you are to plan and organize the solutions, while both of you are to carry out the operations of the trains on your respective boards. Mr. Y, your function is to carry out as well as you can the directions of Mr. X. You are permitted to make suggestions, but the final decision rests with Mr. X. Do you have any questions?

Task

At the start of the first trial, the passenger train was on one siding, pointed so as to move counterclockwise around the track. The freight train was on the main track, pointed so as to move clockwise around the track. The prescribed route for the passenger train was different from the route for the freight, so that the subjects frequently operated the switches to the sidings. Each subject was responsible for the operation of the assigned train. The subjects' task was to complete as many mutually complete trips as possible around the track in a 3-minute trial period. There were six 3-minute trials, with a 1-minute rest period between each trial. The trains were not run simultaneously. The subjects were free to communicate with each other during the entire procedure, but could not ask the experimenter for directions once the first trial had begun.

The achievement score for a given trial was the number of *mutually* complete trips around the track, e.g., if one subject made 8 trips and his partner made 5, the achievement score for that trial was 10. The maximum possible score for a given trial was 32. After the last trial, each subject checked Leary's (1957) Interpersonal Check List (ICL) twice: as he saw himself and as he saw his partner. This 128-item check list can be scored for dominance. The entire procedure lasted one hour.

RESULTS

Achievement

Groups A through G were first compared on age, the Responsibility and Self-Control scales of the CPI, and verbal intelligence. No significant differences between any of the seven groups on any of these variables was found. The means and standard deviations of achievement scores for Groups A through G on Trials 1 through 6 and the sums of trials are presented in Table 3.

Table 4 contains the results of testing Hy-

TABLE 3

MEANS AND STANDARD DEVIATIONS OF ACHIEVEMENT FOR TRIALS 1 THROUGH 6 AND SUMS OF TRIALS

Group		Trials						Sums
		1	2	3	4	5	6	
A	*M*	18.9	23.9	26.8	29.8	30.3	30.8	160.4
	SD	7.0	5.3	5.4	2.1	2.8	2.0	19.0
B	*M*	14.1	19.8	25.7	26.2	26.7	28.7	141.2
	SD	6.7	5.5	4.0	6.0	5.5	4.4	20.1
C	*M*	15.4	23.2	27.4	27.7	28.8	30.8	153.3
	SD	7.5	5.6	4.3	6.5	5.4	1.6	18.7
D	*M*	13.6	20.4	22.5	26.4	29.2	30.8	142.9
	SD	7.7	7.0	8.1	4.5	3.0	2.0	22.8
E	*M*	12.0	20.3	20.7	26.9	26.6	29.2	142.0
	SD	8.7	6.8	3.5	4.2	6.0	3.3	23.0
F	*M*	10.0	18.1	21.2	24.1	27.3	29.5	130.2
	SD	6.9	6.7	4.0	7.8	4.4	2.5	13.8
G	*M*	7.3	14.8	15.9	22.4	27.6	28.4	116.4
	SD	5.7	4.9	7.0	5.1	2.4	4.9	13.5

pothesis 1. The analysis of variance employed was that recommended by Edwards (1950) when there are repeated measurements of the same subjects. Before any of the analyses of variance were performed, Bartlett's test of homogeneity was performed, which proved to be insignificant in all cases.

There is a highly significant difference in achievement between Groups A, B, and G. Inspection of Table 3 shows that Group A was more productive than Group B, which in turn was more productive than Group G. The significant mean square between trials indicates a marked increase in performance across trials for all three groups. The significant interaction of trials by groups reflects the initial higher achievement and greater rapid improvement in performance of Groups A and B over Group G. The analysis of variance confirms Hypothesis 1.

The results of another simple analysis of variance to test Hypothesis 2 are presented in Table 5. The hypothesis was not confirmed, since the mean square between Groups B, E, and F was not significant, and the empirical ordering by achievement was E, B, then F, instead of B, E, then F as predicted. Thus paired dominant subjects (Group E) outperformed paired dominant and submissive subjects (Group B), who in turn outperformed paired submissive subjects (Group F). The mean square between trials was again highly significant, a finding attributable to the increasing achievement with successive trials for all three groups.

A complex analysis of variance was performed to test Hypothesis 3, since the two variables of pairings and role assignments were each varied in two ways. The results are presented in Table 6. The empirical rank order of the four groups was C, D, E, then F, which corresponds to the predicted rank order. The mean squares between pairings and between role assingments were both significant. The results confirm Hypothesis 3. The mean squares between trials was again significant.

TABLE 4
ANALYSIS OF VARIANCE FOR HYPOTHESIS 1

Source	*df*	*MS*	*F*
Between Groups A, B, G	2	811.02	13.91***
Between Trials 1 through 6	5	1,081.78	47.85***
Trials × Groups	10	56.56	2.50**
Between subjects in same group	27	58.31	
Pooled subjects × Trials	135	22.61	
Total	179		

** $p = .01$.
*** $p = .001$.

TABLE 5
ANALYSIS OF VARIANCE FOR HYPOTHESIS 2

Source	*df*	*MS*	*F*
Between Groups B, E, F	2	72.47	1.04
Between Trials 1 through 6	5	1,175.57	41.40***
Trials × Groups	10	19.96	.70
Between subjects in same group	27	69.72	
Pooled subjects × Trials	135		
Total	179		

*** $p = .001$.

Hypothesis 4 predicted the rank order in achievement of all seven groups. The predicted rank order was alphabetical, A through G, while the empirical rank order was A, C, D, E, B, F, and G. The rank order correlation between predicted and empirical rank orders was .78, significant at the .05 level (Table P in Siegel, 1956). This empirical rank order is based on the sums of trials. The rank order correlations, on the other hand, between predicted and empirical rank order calculated separately for Trials 1 through 6 were: .96 ($p = .01$), .64, .61, .75 ($p = .05$), .39, and .55, respectively.

A frequency count of the subjects' spontaneous requests for the maximum possible score shows that 59 of the total N of 70 dominant subjects made such a request, while 12 of the total N of 70 submissive subjects made this request (chi square is 55.6, $p < .001$). Also, 15 additional dominant subjects operated the trains without a partner, as did 15 additional submissive subjects. An F test between the achievement of these two groups was significant at the .05 level, with the dominant subjects achieving more than the submissive subjects. In Mann's (1959) review of the relation

TABLE 6

ANALYSIS OF VARIANCE FOR HYPOTHESIS 3

Source	*df*	*MS*	*F*
Between pairings (Groups C, D vs. E, F)	1	205.35	5.68*
Between roles (Groups C, E vs. D, F)	1	240.00	6.50*
Between Trials 1 through 6	5	1,570.95	43.48***
Trials × Pairings	5	44.95	1.24
Trials × Roles	5	8.88	.25
Pairs × Roles	1	.81	.02
Pairs × Roles × Trials	5	3.66	.10
Within Groups	216	36.13	
Total	239		

* $p = .05$.
*** $p = .001$.

of personality and performance in small groups, dominance is found to be positively related to the number of task contributions.

Perception

Since each subject checked the ICL as he perceived himself and as he perceived his partner, it was possible to determine the subjects' own definition of the situation in terms of dominance of self and partner. The ICL dominance scores are presented in Table 7.

The *intra*group differences in perceived dominance are considered first. The test employed was the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956). Differences that attained statistical significance are presented in Table 8. Inspection of Table 7 shows that for every group which includes a dominant subject, the dominant subjects' self-check mean is always higher than the check of his partner mean. This difference attains statistical significance within Groups A and B. Thus in terms of a comparison of group means, the submissive subjects consistently perceived themselves as less dominant than they perceived their partners and the dominant subjects as more dominant, regardless of the roles assigned and regardless of whether the partner was dominant subject or a submissive subject.

In those groups where a dominant subject and a submissive subject interact (Groups A, B, and G), the dominant subjects' mean check of their partners was always higher than their

TABLE 7

INTERPERSONAL CHECK LIST DOMIANCE SCALE MEANS AND STANDARD DEVIATIONS OF PAIR COMBINATIONS FOR CHECK OF SELF AND CHECK OF PARTNER ($N = 10$ for each pair combination)

Group	Pair Combination	Self		Partner	
		M	*SD*	*M*	*SD*
A	D:d	64.4	3.8	55.8	4.3
	S:s	49.6	9.6	61.2	6.8
B	D:x	62.7	2.9	60.1	6.7
	S:x	49.0	8.5	58.3	7.7
C	D:d	62.9	3.1	60.8	4.7
	D:s	61.7	5.3	61.5	5.9
D	S:d	53.1	8.2	54.3	8.5
	S:s	51.3	7.1	54.1	8.7
E	D:x_1	59.0	3.8	58.9	8.6
	D:x_2	62.6	3.7	57.7	3.8
F	S:x_1	49.8	9.9	57.4	5.9
	S:x_2	49.2	8.1	53.9	8.8
G	D:s	58.0	3.9	56.7	3.5
	S:d	52.1	6.9	57.5	7.0

partners' mean self-checks. This difference attains statistical significance in Group B. In Groups A, B, and G, the submissive subjects' mean self-check was consistently lower than the dominant subjects' mean self-check. This difference is statistically significant in Groups A and B.

Several *inter*group comparisons were made, employing, for the most part, the data as presented in Table 7. The standard deviations of the submissive subjects' self-checks are significantly greater than the standard deviations of the dominant subjects' self-checks ($p < .01$), employing the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956). Thus in their self-checks, dominant subjects are a more homogeneous group than are the submissive subjects. The standard deviations of the dominant subjects' self-checks are significantly greater ($p < .05$) than the standard deviations of the dominant subjects' checks of their partners. Dominant subjects are less uniform in their perceptions of their partners than they are of themselves. This relationship is not true of the submissive subjects.

A comparison of the self-check of the domi-

TABLE 8

STATISTICALLY SIGNIFICANT INTRAGROUP DIFFERENCES IN INTERPERSONAL CHECK LIST DOMINANCE (N = 10 for each pair combination)

Group	Pair Combination with higher Dominance[a]	Pair Combination with lower Dominance[a]	p
A	D:d (s)	S:s (s)	.006
B	D:x (s)	S:x (s)	.006
A	D:d (s)	D:d (p)	.006
E	D:x_1 (s)	D:x_1 (p)	.05
A	S:s (p)	S:s (s)	.01
B	S:x (p)	S:x (s)	.02
B	D:x (p)	S:x (s)	.01

[a] (s) denotes self-check; (p) denotes check of partner.

nant subjects interacting with submissive subjects under three different conditions of role assignment (Groups A, B, and G) shows that the dominant subjects assigned a dominant role (Group A) perceived themselves as more dominant than the dominant subjects assigned no role (Group B), who in turn perceived themselves as more dominant than the dominant subjects assigned a submissive role (Group G). These three groups are significantly different ($p < .02$), employing the Friedman two-way analysis of variance (Siegel, 1956). A similar (statistically insignificant) rank order holds for the submissive subjects' check of their respective dominant partners in Groups A, B, and G, i.e., the submissive subjects in Group A rated their own dominant partners as more dominant than did the submissive subjects in Group B rate their own partners. The submissive subjects in Group B in turn rated their dominant partners as more dominant than the submissive subjects in Group G rated their own dominant partners.

DISCUSSION

The general pattern of the achievement of the seven groups tends to validate the hypotheses, both in statistical significance of the differences between groups and in the direction and ranking of the predicted differences. The failure of the results to support Hypothesis 2 is due largely to the unpredicted high achievement of Group E. Group C also achieved higher scores than predicted in Hypothesis 4. Both of these groups have paired dominant subjects. The greater achievement of these two groups is in part due to the higher aspiration level of the dominant subjects, as previously noted.

With the exception of Group G, the assignment of dominant and submissive roles to any of the pairings resulted in an increase in achievement. The greatest absolute increase in achievement as a result of assignment of roles was between the dominant subject with a submissive subject pairings (Group A compared with Group B). The assignment of roles to paired dominant subjects or paired submissive subjects resulted in virtually the same absolute increase in achievement, i.e., the difference between Group C and Group E is similar to the difference between Group D and Group F. The assignment of roles gave a structure within which the subjects were able to function more effectively than the comparable pairings without role assignment.

There was a rapid rise in achievement across trials for all groups, so that by Trial 6 all groups were performing at virtually the same level. This similarity of level of performance by Trial 6 is in part an artifact of the task, since the top possible score for a trial was 32. Group G made the greatest absolute gain from Trial 1 through Trial 6. However, this is due in part to the low initial performance on Trial 1. If there had been, e.g., only three trials instead of six, the differences in achievement would have been greater and in the same direction. The major differences in achievement between groups appeared in the initial trials. The effect of the relative dominance of the partners as well as the assignment of roles is greatest here.

The ICL results indicate that the subjects' definition of the situation in terms of dominance closely reflected both the selection of the subjects high and low in personal dominance as well as the assignment of dominant and submissive roles. The degree of dominance ascribed to the partner, relative to the subjects' own dominance, is a function of the subjects' personal dominance. These findings are in accord with those of Leary (1957) and Naboisek (1953). The degree of dominance ascribed to the self and the partner varied in the direction of the assigned role.

These relationships between the subject's own dominance and the degree of dominance he ascribes to others are in contrast to a study

(Altrocchi, 1959) employing the same pool of subjects as the present study, where the dominant subjects did not ascribe less dominance to others than the submissive subjects. One critical difference between these studies is that Altrocchi's subjects observed movies of social objects before rating them, whereas the subjects in the present study interacted before rating each other, thus both communicating as well as evoking cues (Bruner & Tagiuri, 1954) of dominance or submissiveness.

Both the dominant and submissive subjects attributed considerable dominance to their partners, regardless of the partner's personal dominance. This finding may be attributed in part to the cooperative nature of the task as well as the fact that dominance is an approved trait for males, so that attributing dominance to another is a way of indicating social desirability (Edwards, 1957).

Generalizations from the achievement and perceptual findings are restricted in that the subjects were males chosen from a rather restricted socioeconomic-educational stratum of society as well as from the extremes of the distribution of dominance. However, the task is representative (Brunswik, 1947) of many life experiences, where two individuals work together in the pursuit of some objectively mutual goal under circumstances requiring a good deal of communication and cooperation.

SUMMARY

This study related different pairings of dominant and submissive males, who interacted under various role assignment conditions, to their joint achievement in a cooperative problem solving situation. Predictions of differential achievement were derived from the thesis that the productivity of a pair depends upon the degree to which conditions permit each member to utilize his habitual patterns of interpersonal behavior.

A personality inventory was administered to 748 male university undergraduates. Volunteer subjects scoring very high or very low on a dominance scale were selected for participation in the cooperative task situation. The task was the operation of two model railroad trains for a series of six 3-minute trials. All groups were homogeneous with respect to age, and score on social responsibility, self-control, and vocabulary tests.

Of four predictions concerning the relative achievement of the seven interaction groups, three were confirmed. The most productive group was composed of pairs in which the dominant subject was assigned the dominant role and the submissive partner the submissive role. The least productive group was composed of these pairings with the roles reversed. Paired dominant subjects were more successful than paired submissive subjects, and both of these pairings achieved more when assigned roles. All groups showed significant increases in performance across trials.

The degree of dominance attributed to others was a function of the subjects' personal dominance. Dominant subjects perceived their partners as less dominant than themselves and submissive subjects perceived their partners as more dominant than themselves. The degree of dominance ascribed to the self and the partner varied in the direction of the assigned role.

It was concluded that congruence of role and habitual pattern within the subject and complementarity of patterns as between subjects were major determining variables in cooperative achievement. Sullivan's (1953) personality theory was employed in theoretical consideration of these results.

REFERENCES

ALTROCCHI, J. C. Dominance as a factor in interpersonal choice and perception. *J. abnorm. soc. Psychol.*, 1959, **3**, 303–308.

BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 17.

BRUNSWIK, E. *Systematic and representative design of psychological experiments, with results in physical and social perception*. Berkeley: Univer. California Press, 1947.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

EDWARDS, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *J. abnorm. soc. Psychol.*, 1957, **55**, 394–396.

GHISELLI, E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *J. abnorm. soc. Psychol.*, 1958, **57**, 61–66.

GOUGH, H. G. *Manual for the California Psychological*

Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1957.

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

MANN, R. D. A review of the relationships between personality and performance in small groups. *Psychol. Bull.*, 1959, **56**, 241–270.

NABOISEK, H. Validation of a method for predicting role expectations in group therapy. Unpublished doctoral dissertation, University of California, 1953.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.

THORNDIKE, R. L., & GALLUP, G. H. Verbal intelligence of the American adult. *J. genet. Psychol.*, 1944, **30**, 75–85.

TUMA, A. H. An exploration of certain methodological and client-counselor personality characteristics as determinants of learning in the counseling of college students. Unpublished doctoral dissertation, University of Maryland, 1955.

(Received April 20, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 543-552

DELAY OF GRATIFICATION, NEED FOR ACHIEVEMENT, AND ACQUIESCENCE IN ANOTHER CULTURE[1]

WALTER MISCHEL
Harvard University

RELATIONSHIPS between overt behavior on important psychological dimensions and relevant motives are much sought foundations in personality research. This paper explores the relationship between a person's choice preferences for immediate, smaller (ImR) as opposed to delayed, larger (DelR) reinforcement and his motive system. It thus seeks to extend our knowledge of preference for delayed gratification, some antecedents and correlates of which have already been investigated (Mischel, 1958, 1959, 1961). Here choice behavior with respect to preference for ImR or DelR is related to the need for achievement (n Achievement) as scored from fantasy material in the manner described by McClelland and his associates (McClelland, Atkinson, Clark, & Lowell, 1953).

The central aim is not primarily to elaborate the network of correlates of n Achievement, but rather, to clarify some of the variables relevant to preferences for ImR or DelR by relating such choices to a motive that is logically relevant and whose measurement and empirical correlates appear relatively firm. If this can be done, research on delay of gratification with human subjects and in choice situations, in which smaller, immediate gratifications are pitted against more long term, larger gratifications, should be somewhat more solidly anchored, albeit only on a correlational basis, to the data of other more elaborated domains of personality research.

[1] Thanks are due to David C. McClelland and to Ralph Metzner for informative pre- and post-experimental discussions on some of the central problems dealt with in this paper, to John W. Whiting for important suggestions concerning the measures used and for his "magic man" procedure, and to Claire Gerschuni and Judith Vollmar for statistical computations. Grateful acknowledgment is also made to the Education Office of the Government of Trinidad and Tobago, and its officials for cooperation in testing within their schools. This study was carried out under the partial support of a grant from the Laboratory of Social Relations, Harvard University, and Grant M-2557(A) from the National Institute of Mental Health.

The ability to delay immediate gratification for the sake of later, more gratifying outcomes has long been recognized as crucial in a multitude of complex human situations; it figures importantly in a variety of theoretical formulations (e.g., Freud, 1922; Rapaport, 1951; Singer, 1955). In spite of this recognition, there is a relative paucity of relevant experimental research. Recently, however, a number of investigations (e.g., Levine, Glass, & Meltzoff, 1957; Levine, Spivack, & Wright, 1959; Singer, Wilensky, & McCraven, 1956; Spivack, Levine, & Sprigle, 1959) have inquired into the correlates of "delaying capacity," primarily as inferred from the frequency of human movement (*M*) Rorschach responses. In contrast to this line of research, the present study is part of a larger research program that approaches the problem of delay of gratification by using direct measures of preference for ImR or DelR in particular choice situations. Thus far, with children from other cultures as subjects, preference for DelR as opposed to ImR has been found to relate positively to social responsibility and to accuracy of time statements, negatively to delinquency (Mischel, 1961), positively to intelligence (Melikian, 1959), and positively to the presence (as opposed to the absence) of the father within the home under some conditions (Mischel, 1958, 1959). The method also has been used to test cross-cultural hypotheses experimentally with respect to differences in delay of gratification (Mischel, 1959).

Research on the need for achievement requires no new summary (e.g., Atkinson, 1958). It is consistent with past theorizing and research on n Achievement that high n Achievement should be related to preference for DelR as opposed to ImR. The essence of the definition of n Achievement is "competition with a standard of excellence" (McClelland et al., 1953). Implied in such striving is the ability to delay and postpone the relatively trivial but immediately available for the sake

of later but more important outcomes. Achievement fantasies may be thought of in part as reflecting as well as sustaining and mediating the individual's strivings for future rewards and attainments of excellence. Indeed, imbedded in the operations for scoring n Achievement is the criterion of "long-term involvement."[2] Clearly there can be little realization (and thus little maintenance) of the motive to compete with a standard of excellence unless the person is able to delay immediate but smaller gratifications and to choose instead larger future rewards and goals.

A direct and positive relationship between preference for DelR and n Achievement, then, is to be expected for the following reasons. It has been argued (Metzner, 1960) that one of the crucial conditions that facilitates the development of the ability to delay gratification is the acquired reward value of working itself. Learning to delay starts when behaviors based only on the pleasure principle are either frustrated or punished. Then, fantasy becomes an acquired primary process substitute and work a secondary process detour, both of which are rewarded by eventual (larger) gratifications, thus ultimately becoming rewarding activities in their own right. Liking to work for its own sake is generally assumed to be a basic ingredient of the high n Achievement pattern. Presumably, persons high in n Achievement have learned to like work, and they have learned this in part as a response to demands to forego immediate gratification in favor of more long-term goals. That is, they have learned to tolerate waiting periods and in the course of such socialization have built up a general readiness to delay when it is demanded by conditions or people. Consequently, when persons high in n Achievement are given choices between immediate, smaller rewards and delayed, larger ones, the balance between the negative valence of waiting and the positive valence of the delayed, larger reward is weighted in favor of the latter because the former is relatively minor for them. This is in contrast to the state of affairs with persons who have not learned work as a response to the delay situation, or who have not acquired intrinsic interest in achievement (i.e., persons low in n Achievement). The anticipated positive relationship between preference for DelR and n Achievement does not depend on an absolute preference by individuals high in n Achievement for the delayed alternative, but, rather, applies equally to relatively less nonavoidance of the delayed alternative on their part.

In addition to the central aim of testing the relationship between the need for achievement and patterns of preference for reinforcement with respect to delay of gratification, the paper also presents data relevant to the measurement and correlates of n Achievement itself, and to further correlates of preference for ImR or DelR. Especially for use in cross-cultural research, in which verbalizations in response to Thematic Apperception Test (TAT) pictures appropriate for the usual scoring of n Achievement often are difficult or impossible to obtain, alternative n Achievement measures are much needed. Graphic materials ("doodles") scored for n Achievement (Aronson, 1958) have been used with young children. The present study offers another alternative to measuring n Achievement via the TAT, far simpler than the Test of Insight (French, 1958), and possibly appropriate for use in many cultures.

Further, an attempt was made to use a measure—suitable for research in another culture—of simple acquiescence or "agreeing tendency." Much recent research with United States samples (e.g., Bass, 1956; Messick & Jackson, 1957) has focused on the response set to agree, and an intensive investigation of the problem (Couch & Keniston, 1960) has related "yeasaying" as opposed to "naysaying" to two personality syndromes. These syndromes are conceptually germane to the present distinction concerning preference for ImR-DelR. "Yeasayers were shown to be individuals with weak ego controls, who accept impulses without reservation, and . . . easily respond to stimuli exerted on them. The naysayer inhibits and suppresses his impulses . . ." (p. 173). The

[2] Since "long-term involvement" is a criterion used in scoring for the presence of Achievement Imagery, and may have direct overlap with the ability to delay reinforcement, it is important to note that on only seven of the protocols did this criterion serve as the basis for Achievement Imagery scoring, and three of these protocols were given by subjects not consistently preferring delayed, larger reinforcement on the choice measures.

former "can freely indulge in impulse gratification" whereas "the egos of naysayers take over the controlling functions of the parents, and the suppression of impulses is subsequently self-maintained and self-rewarded" (p. 173). A cluster of scales including impulsivity and dependency characterized the positive end of the agreeing tendency; the disagreeing tendency was defined by scales including ego strength, responsibility, and trust. Further, in interviews, yeasayers (in contrast to naysayers) described themselves as ". . . unreflective, quick to act, easily influenced, and unable to tolerate delays in gratification" (p. 170).

Clearly, these two personality syndromes are directly relevant to preference for ImR-DelR. Indeed, I have reasoned previously (Mischel, 1958), with some direct experimental support (Mahrer, 1956), that "trust," or the expectancy that the promised delayed reward is forthcoming, is a basic factor in preference for DelR, and empirically have related preference for DelR to social responsibility (Mischel, 1961). It is to be expected that DelR preference should be related to naysaying and ImR preference to yeasaying and to the personality syndromes associated with each. Similarly, in the realm of n Achievement research, numerous attempts, mostly with mixed or dubious success (e.g., Samelson, 1958) have been made to relate n Achievement inversely to conforming behavior. Acquiescence may be thought of as a special case of conforming in an ambiguous situation, and again in the present research the often expected but rarely attained relationship between n Achievement and nonacquiescence was anticipated.

Lastly, data were also collected on social responsibility, as measured by an independently validated scale (Harris, 1957), substantially correlated with other measures of adjustment and maturity. Replication was thus sought for a previously established positive relationship between measures of preference for DelR (as opposed to ImR) and social responsibility (Mischel, 1961) and to determine the extent of interrelationship between social responsibility and the other variables measured in this study.

The over-all methodological aim is to aid the development of relatively simple but quantifiable techniques for testing psychological hypotheses in cultures much different from our own in which measures developed in the United States are inappropriate or excessively cumbersome. Such techniques are needed if the generality of psychological principles established on United States samples is to be extended to grossly different populations.

METHOD

Subjects

A total of 112 Trinidadian Negro children (68 boys and 44 girls), all in the age group 11–14, with a modal age of 13, were tested in a government school located near the capital city. The school contained children from the urban environs, primarily from the lower middle and lower socioeconomic strata of the society.

Measures of Preference for ImR or DelR

Three preference measures (a behavioral choice and two questionnaire items) were used to elicit preference for ImR or DelR. These were selected on the basis of earlier research with comparable Trinidadian samples. The general procedures used in the pretesting and selection have been described in detail (Mischel, 1959, 1961). Briefly, the behavioral measure consisted of a choice between ImR or DelR in the form of a small candy bar (costing approximately 10 cents) available immediately *or* a much larger candy bar (costing approximately 25 cents) for which the subject must wait one week.[3] The reward choice was administered as the last item in the total procedure and was structured as a way of "thanking" the subjects for their cooperation. The two verbal items, one inserted near the beginning, the other near the middle of the procedure, were:

1. I would rather get ten dollars right now than have to wait a whole month and get thirty dollars then.
2. I would rather wait to get a much larger gift much later rather than get a smaller one now.

Previous research (Mischel, 1961) has shown that the subject's responses over three such measures are significantly interrelated and that the three measures in combination tend to provide a more useful classification of subjects with respect to ImR or DelR preferences than does a single measure. Consequently, following the 1961 procedures, subjects were divided for purposes of data analysis into "Consistent ImR" (ImR preference over the three measures), "Consistent DelR" (consistent DelR preferences over the three measures), and "Inconsistent R" (one ImR and two DelR responses, or the reverse). Although the predictions center on differences between the Consistent ImR and Consistent DelR groups, results are also presented in terms of the number of DelR choices out of the total three

[3] The size of the immediately available reinforcement used here was deliberately greater than that used in past work (Mischel, 1961) in order to increase the probability of a more equal dichotomy in the immediate, smaller versus delayed, larger choice distribution.

possible made by the subject, i.e., 3 DelR, 2 DelR, 1 DelR, or 0 DelR.

Measures of n Achievement

Two measures of n Achievement were used. The first was the standard group-administered procedure developed by McClelland and his associates (1953), using five TAT-type cards.[4] The second measure was an innovation, in the form of an open-ended aspiration question. The latter was introduced in the context of "let's pretend there is a magic man": "Now let's pretend that the magic man who came along could change you into anything that you wanted to be, what would you want to be?" Subjects were instructed to answer in one word.

Measure of Acquiescence

A simple measure of acquiescence or conformity suitable for large scale research in other cultures—simple to answer, simple and rapid in administration, and meaningful in other cultural contexts—was sought. The measure adopted consisted of the following instructions, administered orally by the experimenter:

> I have something in mind. I am closing my eyes and concentrating and thinking of it . . . it is something that you might agree with or that you might disagree with . . . you might disagree with what I'm thinking of or you might agree with it . . . if you agree put down a *Y* for Yes; if you disagree put down a *N* for No. (This was repeated twice, reversing the position of *Y* and *N* in the sequence of instructions.) Now I'm concentrating on it (closing eyes); go ahead.

Measure of Social Responsibility

Social responsibility was determined by response to a quite carefully constructed and independently validated Social Responsibility Scale (Harris, 1957). The Social Responsibility Scale (SRS) used has been found to correlate substantially with other measures of personal and social adjustment and was designed to "discriminate children who have, with their peers, a reputation for responsibility as contrasted with children who have little reputation for responsibility" (p. 326). All items from the 50-item SRS (Harris, 1957, called "Social Attitudes Scale," p. 324, Table 1), with the exception of Item 50, were administered in the modified form described earlier (Mischel, 1961).

Procedure

The experimenter was introduced as an American from a college in the United States interested in gathering information on the children in the various schools of the island. To help with this the subjects were asked to answer a number of questions. The details of the instructions are similar to those previously reported (Mischel, 1961). Subjects were seated sufficiently far apart from each other to give reasonable assurance that their choices were made independently. All testing was done in group settings in large classrooms, with two sessions. All measures except the TAT and the candy choice were given in the first session.

RESULTS AND DISCUSSION

Examination of both sex and age in relation to preference for DelR or ImR revealed no differences approaching statistical significance and consequently the male and female data and all age groups were treated collectively in subsequent analyses.

n Achievement and Preference for Delayed Reinforcement

Inspection of the responses to the five TAT-type cards indicated that the protocols were, on the whole, impoverished and short, with relatively little elaboration, a finding not unexpected for this culture. Simplified scoring for n Achievement was therefore decided upon. Protocols were scored only for presence or absence of achievement imagery (AI or UI, scored +1 or −1), or doubtful achievement imagery (TI, scored 0), in the manner described by McClelland et al. (1953), without continuing additional scoring of AI cards. This simplified procedure avoided the necessity for making subtler and potentially less reliable scoring decisions on relatively fragmentary responses that would be especially difficult to interpret, as normative scoring samples are not available for the culture. Since the range of scores obtained with the shortened scoring method for five cards extended from +5 to −5, and since any lack of discrimination through simplified scoring is likely to result in inability to reject the null hypothesis, the procedure seemed justifiable.

Comparison of subjects with respect to delay of reinforcement patterns as they related to n Achievement was done as follows. Subjects were first divided into Consistent ImR, Consistent DelR, and Inconsistent R, to categorize their choices with respect to the three measures of preference for ImR or DelR, the choice preferences with respect to DelR also being shown in terms of the number of DelR choices

[4] The five TAT-type pictures used were two men ("inventors") in a shop working at a machine; a boy in a checked shirt at a desk, an open book in front of him; a man at his desk in a dark office; a man looking at photos; a family at dinner. They were administered in printed booklets with the standard four questions printed for each picture on a separate page. The only modification made in the reproduction of the cards was a darkening of the pigmentation of the people in the pictures.

TABLE 1
REINFORCEMENT PREFERENCE WITH RESPECT TO DELAY IN RELATION TO n ACHIEVEMENT SCORES, ASPIRATIONS, AND ACQUIESCENCE

Reinforcement Preference: *N* DelR[a]	n Achievement			n Achievement Distribution of Scores[b]					Aspirations[c]			Acquiescence	
	N	*M*[e]	*SD*	−5 to −3	−2 to −1	0	+1 to +2	+3 to +5	Occupational	Achievement Trait	Personal Trait	Yes	No
Consistently Delayed: 3	37	.92	2.35	1	12	6	7	11	18	4	14	25	12
Inconsistent: 2	18	1.11	2.32	0	4	4	6	4	8	2	7	12	6
1	27	0 −.74		8	4	6	8	1	4	5	17	21	6
Consistently Immediate: 0	30	−.70	2.62	11	6	3	5	5	9	0	17	27	3
Totals:	112			20	26	19	26	21	39	11	55	85	27

[a] *N* of delayed reward choices made by subject.
[b] Subject's total score for the five TAT cards: *N* of subjects in each category of distribution.
[c] *N* of subjects giving Occupational aspirations, Achievement Trait aspirations, and Personal Trait aspirations; seven subjects failed to state any aspirations.
[e] *t* of 3 DelR versus 0 DelR on mean n Achievement = 2.61, $p < .01$; *t* of 3 DelR + 2 DelR versus 1 DelR + 0 DelR on mean n Achievement = 3.78, $p < .001$.

made by the subject out of the three total possibilities. Thus the Consistent ImR group contains all subjects with 0 DelR choices, the Inconsistent group is further subdivided into subjects with one as opposed to subjects with 2 DelR choices, and the Consistent DelR group contains all subjects with 3 DelR choices.

Table 1 presents the means and standard deviations of n Achievement scores for subjects in each of these categories with respect to reinforcement preference. As anticipated, the mean n Achievement score of Consistent DelR subjects as compared to Consistent ImR subjects is significantly higher ($t = 2.61$; $p < .01$). (One-tailed tests of significance are used throughout this paper.) The hypothesized relationship between preference for ImR as opposed to DelR and n Achievement is thus strongly supported. Similarly, if subjects are dichotomized into "High DelR" (2 or 3 DelR responses, $M = .98$; $SD = 2.22$), and compared on mean n Achievement with "Low DelR" (0 or 1 DelR response, $M = -.41$; $SD = 2.46$), the resulting t is 3.78 ($df = 110$; $p < .001$). Clearly, subjects showing greater preference for DelR have significantly higher n Achievement scores than subjects with lesser preference for DelR. Of secondary interest is the fact that the mean n Achievement scores of subjects in the Inconsistent R category are between those of the Consistent DelR and Consistent ImR groups. The difference in means between the Consistent DelR group and the Inconsistent R group reaches significance ($p < .05$), but the difference in means between the latter and the Consistent ImR group does not reach an acceptable level. If the data are analyzed in correlational terms, a point biserial correlation of .31 ($p < .01$) is obtained from the correlation of n Achievement scores with Consistent DelR or Consistent ImR choice. A Pearson correlation between the number of DelR choices (0–3) and n Achievement scores yields an r of .27 ($p < .005$).

Responses to the open-ended aspiration question used to measure n Achievement were classified into three categories that emerged from inspection of the data: "Occupational" responses and "Trait" responses, the latter being further subdivided into "Achievement Traits" and "Personal Traits." The first category contained all responses mentioning an occupation or profession (e.g., policeman, pilot, doctor, priest); the second contained all responses mentioning personal traits that appeared to be directly achievement-related (e.g., important, big shot, bright, successful); the third contained all other responses, and consisted of traits that were not explicitly achievement-related (e.g., nice, same, a baby,

small, adult, brother, honest, kind, brown). Judgments of whether a response belonged in the second or third category were made by two raters with knowledge of the culture; ratings showed more than 90% agreement. The Occupational category was of course extended to include all possible occupational aspirations within the culture. Thus, for example, king, prince, queen were included in the first category since Trinidad is within the British Commonwealth and such aspirations are analogous to "president" in a United States sample. The major conceptual distinction between the aspirations in the first as opposed to the second and third categories is that the former are clearly long-term aspirations of a career type, whereas the latter are not necessarily long-term goals, and are not explicitly career goals, but rather personal qualities.[5]

Numerous other distinctions can of course be made in scoring such responses. The major reason for using the present categorization is that when Occupational and Achievement Trait categories were combined it classified responses into a dichotomy which is as close to a 50-50 division as it was possible to obtain and which also appears to draw, albeit grossly, the desired conceptual distinction. The alternative possible distinction between occupational responses and all others is firmer since it is not dependent on nearly as subtle a judgment as that entailed in discriminating between kinds of traits. Statistical comparisons were made in which the first category was opposed to the other two combined, as well as in which the first two categories were combined and opposed to the third.

To determine the usefulness of this open-ended aspiration question for measuring n Achievement, responses in each category were related to the mean of n Achievement scores as measured in response to the more standard TAT-type procedure. Table 2 presents the mean and standard deviations of n Achievement scores for subjects giving aspiration responses in each of the three categories and in combined categories, as well as the number of subjects in various categorizations of the

[5] The crudeness of the present scheme should be clear. For example, "policeman," although an occupational aspiration, may be less related to striving for achievement than is the aspiration to be "honest."

TABLE 2

n ACHIEVEMENT IN RELATION TO ASPIRATIONS AND ACQUIESCENCE

n Achievement scores[a]	*N* of subjects giving each kind of aspiration			Acquiescence	
	Occupational	Achievement Traits	Personal Traits	Yes	No
Positive					
+3 to +5	13	2	6	15	6
+1 to +2	12	5	10	17	9
0	6	1	9	14	5
Negative					
−1 to −2	7	3	14	21	5
−3 to −5	1	0	16	18	2
Totals:	39	11	55	85	27
n Achievement *M*:	1.44	1.09	−.73	.04	.79
	1.36		−.42		
n Achievement *SD*:	1.89	2.15	2.56	2.53	2.25
	1.96		2.59		

Note.—*t* of Occupational aspirations versus all other aspirations on mean n Achievement = 3.88; $p < .001$. *t* of Occupational aspirations + Achievement Trait aspirations versus Personal Trait aspirations on mean n Achievement = 4.54; $p < .001$.

[a] Total score for the five TAT cards.

n Achievement distribution giving each kind of aspiration. The difference between the n Achievement means of subjects giving Occupational responses as compared to all others is significant ($t = 3.88$; $p < .001$), the mean of n Achievement scores of subjects in the group giving occupational aspirations being significantly higher. Similarly, comparison of the mean n Achievement of subjects giving occupational responses or Achievement Traits versus all other responses results in a *t* of 4.54 ($p < .001$), the former having significantly higher n Achievement scores than the latter. A Pearson correlation between Occupational responses, Achievement Trait responses, and Personal Trait responses (assigned values from 2 to 0) and n Achievement scores yields an *r* of .41 ($p < .001$).

This short-cut method may serve as a useful alternative to the usual much longer procedure of scoring fantasy material for n Achievement, especially when the latter is difficult to obtain, as it is in many kinds of

field work. Some very recent and independently formulated attempts to relate n Achievement to realism in vocational aspirations (Mahone, 1960) should be noted, and provide further encouragement for exploring relationships between motive variables and aspirations.

The data for the relationship between aspirations and reinforcement preference patterns with respect to delay are presented in Table 1. A chi square comparison of subjects giving 3, 2, 1, and 0 DelR choices with the number of responses given by these groups in each of the three aspiration categories cannot be made meaningfully since this results in an excess of expected frequencies that are less than 5. Combining categories, subjects giving 2 or 3 DelR choices ("High DelR") were compared with subjects giving 0 or 1 DelR choices ("Low DelR") in regard to the number of responses in each of the three categories of aspirations. This comparison results in a chi square of 7.49 ($df = 2$; $p < .05$). Subjects with "High" DelR preference tend to give more Occupational and fewer Personal Trait responses than subjects with "Low" DelR preference. Chi square analyses of High-Low DelR in relation to number of Occupational aspiration responses versus number of all other aspiration responses, or in relation to a combination of the Occupational category and the Achievement Trait category as opposed to the Personal Trait category, both show a significant relationship of the same nature ($p < .02$, chi squares corrected for continuity).

Acquiescence, Reinforcement Preference, and n Achievement

The acquiescence data for yeasaying or naysaying in an ambiguous situation are related to the reinforcement preference patterns in Table 1. Consistent DelR subjects gave 25 yes and 12 no responses; consistent ImR subjects gave 27 yes and 3 no responses. Application of the chi square test to these data results in an uncorrected chi square of 4.80 ($df = 1$; $p < .05$). Similarly, a dichotomization into "High DelR" as opposed to "Low DelR" to compare number of yes or no responses, results in an uncorrected chi square of 4.38 ($df = 1$; $p < .05$).[6] As anticipated, subjects high in preference for DelR gave relatively more no and fewer yes responses than did subjects low in preference for DelR.

In spite of its relative weakness, this association is striking since only one item was used to measure acquiescent yeasaying tendencies. The relationship between yeasaying and naysaying tendencies and reinforcement preference patterns would seem well worth further study with more reliable multiple measures to obtain yeasaying data. The present measure of acquiescence may have advantages in studying various aspects of acquiescent behavior by permitting easy manipulation of the reference group, e.g., merely changing the public identification of the examiner. The method may also be used readily to study conforming tendencies cross-culturally. The preponderance of yes responses in the present results may be taken as a caution against excessive generalizations about strong conforming tendencies in our own culture without reference to comparative data. The primary relevance of these data, however, is that they serve as an empirical link between the patterns isolated by Couch and Keniston (1960) to distinguish yeasayers and naysayers and the present distinction between subjects consistently preferring DelR or ImR. Clearly the patterns have strong conceptual overlap: the present findings provide support for their empirical overlap.

The n Achievement scores obtained both from the standard n Achievement measure and the aspiration measure were related to yeasaying as opposed to naysaying. Theoretical considerations about the personality correlates of high n Achievement lead to the anticipation that persons high in n Achievement should show less acquiescence or yeasaying. However, although the mean n Achievement scores (on the standard measure) of nonacquiescent subjects is higher ($N = 85$; $M = .04$; $SD = 2.53$) the difference clearly falls short of acceptable significance levels and can only be interpreted as a trend ($t = 1.35$;

[6] If extreme caution is used and the decision is made to apply the Yates correction for continuity on all 2×2 contingency tables, although all expected frequencies in each cell do exceed 5, the comparison of Consistent DelR with Consistent ImR subjects on acquiescence results in a chi square of 3.59; the comparison of the "High" versus "Low" preference for DelR dichotomy with acquiescence results in a chi square of 3.51. A chi square of 3.84 is required for $p < .05$ and of 2.71 for $p < .10$.

$.10 > p > .05$, one-tailed test). Table 2 shows the number of yes and no responses given to the acquiescence measure by subjects in various categorizations of the total distribution of n Achievement scores for the five TAT stories. Comparison of subjects with positive scores as opposed to subjects with negative scores on the number of yes versus no responses results in an uncorrected chi square of 3.59 which approaches but does not reach an acceptable significance level ($p < .10$; chi square corrected for continuity = 2.72, $p < .10$). If subjects with zero and negative n Achievement scores are combined and compared with subjects having positive n Achievement scores, this represents the closest median split possible, using all subjects on the yes-no dichotomy. An uncorrected chi square of 2.73 with $p < .10$ results. In sum, a trend of such dubious statistical significance appears to characterize the relationship between our measure of acquiescence and the TAT n Achievement measure that we cannot safely reject the null hypothesis. As such, the present findings share the uncertain conclusions about the relationship between acquiescence and n Achievement obtained by other investigators (e.g., Samelson, 1958).

The data also permit a test of the relationship between the aspiration measure of n Achievement and acquiescence (Table 3). If subjects giving Occupational aspirations are combined with those giving Achievement Trait aspirations and compared to those giving Personal Trait aspirations with respect to the frequency of the yes or no responses, the resulting chi square does not approach statistical significance. However, comparison of subjects giving Occupational aspirations with those giving all other aspirations on acquiescence results in a chi square of 4.27 (corrected for continuity), $p < .05$. Subjects giving Occupational aspirations, when compared to those giving all other aspirations, significantly show relatively less yeasaying and more naysaying. It is of interest that the aspiration measure of n Achievement is thus significantly related to acquiescence whereas results from the more standard fantasy measure of n Achievement fall short of statistical significance. Even more clear-cut relationships may perhaps be expected when data that lend themselves to more detailed and subtle scoring are employed and when multiple acquiescence measures are used.

TABLE 3
ACQUIESCENCE IN RELATION TO EACH CATEGORY OF ASPIRATIONS

Acquiescence	Aspirations		
	Occupational	Achievement Traits	Personal Traits
Yes	24	10	44
No	15	1	11

Note.—Chi square comparison of Occupational aspiration; with all others on Yes versus No: $\chi^2 = 4.27$ (corrected for continuity); $p < .05$.

Social Responsibility in Relation to the Other Measures

As regards the SRS, the relationship between preference for DelR and social responsibility reported earlier (Mischel, 1961) is replicated. A Pearson correlation of .28 ($p < .005$) is obtained between number of DelR choices (0–3) and SRS scores. Using t tests, significant differences between means ($p < .01$) are obtained when "High DelR" subjects (3 and 2 DelR choices) are compared with "Low DelR" subjects (1 and 0 DelR choices). The means and *SD*s of the former are 31.25 and 4.28, and of the latter 29.05 and 3.70. A secondary finding of considerable interest (Table 4) is a strong relationship between SRS and n Achievement ($r = .40$, $p < .001$ with TAT measure; $r = .33$, $p < .001$ with aspiration measure). This appears to be one of the very few instances in which a significant positive correlation is obtained between n Achievement and a *questionnaire* measure. Theoretically, although not specifically predicted, it seems extremely plausible that social responsibility and achievement motivation should be interrelated. Operationally, it is encouraging that at least in this cultural setting, questionnaire data do appear correlated with scores based on fantasy material.

An Overview

Summarizing the results, Table 4 presents the intercorrelations of preference for DelR, the two n Achievement measures, and SRS. Preference for DelR, n Achievement measured by two different but intercorrelated methods (TAT and open-ended aspirations), and SRS scores are all significantly and positively

interrelated. The acquiescence measure is not significantly associated with social responsibility scores ($r = .06$, $p < .20$), although, as discussed above, low significant relationships do seem to hold between naysaying and preference for DelR and between naysaying and Occupational (as opposed to all other) aspirations. Finally, partial correlations were computed between preference for DelR and n Achievement, holding SRS constant, and between preference for DelR and SRS, holding n Achievement constant. In both cases, low but positive and still significant correlations result ($r = .18$ and .19, respectively, $p < .05$). This finding suggests that the major measures used to delineate the cluster of correlates associated with preference for DelR as opposed to ImR, while themselves significantly intercorrelated, still appear to tap different aspects of the constellation.

The main import of the over-all findings is that they relate reinforcement preference to other more thoroughly explored variables. When the present findings are taken collectively with those obtained on other previously reported samples, a network of meaningful correlates becomes apparent. These are interpreted as providing considerable evidence that an obviously crucial but relatively unstudied dimension of behavior, the ability to delay the relatively trivial but immediately available for the sake of larger but later outcomes, can be explored usefully with relatively simple direct measures of choice preference. The correlates of these choice preferences are becoming increasingly clear, although controlled studies of antecedents are still lacking.

The simple preference patterns with respect to delay elicited thus far provide only an early model and a prelude to more complex and subtler batteries. In such batteries, delay preference patterns systematically sampling a variety of value and need areas could be tapped, and the valences of particular intervening work and fantasy activities (in addition to the simple waiting period over a time interval used at present) could be elicited. The ultimate aim would be the construction of predictive individual topographies for complex choices on the dimension of delay, and the measurement of individual time lines.

The findings also support the fruitfulness of devising simple quantifiable measures for eliciting data on potentially complex dimensions in field research situations in other cultures. Although such field research is often thought to present serious methodological problems when quantifiable data on psychological dimensions are sought, the present results indicated, on the contrary, that exceedingly simple measures may serve to yield useful data that might only be elicited from more sophisticated subjects by much more complex methods. In the present situation, the subjects' lack of psychological sophistication appears to have been a strong asset in circumventing sources of error that so often bedevil comparable researches.

TABLE 4
INTERCORRELATIONS OF PREFERENCE FOR DELAYED, LARGER REINFORCEMENT (DelR), n ACHIEVEMENT, AND SOCIAL RESPONSIBILITY (SRS)
($N = 112$)

	n Achievement		SRS
	TAT	Aspirations[a]	
Preference for DelR	$r = .27$**	C = .26*	$r = .28$**
n Achievement			
TAT		$r = .41$***	$r = .40$***
Aspirations[a]			$r = .33$***

[a] $N = 105$.
* $p < .05$ ($\chi^2 = 7.49$, $df = 2$).
** $p < .005$.
*** $p < .001$.

SUMMARY

This study, using 112 Trinidadian subjects aged 11–14, tested the relationship between preference for immediate, smaller (ImR) or delayed, larger (DelR) reinforcement in choice situations and (*a*) n Achievement, (*b*) acquiescence (yeasaying or naysaying in an ambiguous situation). The expected significant positive relationships between preference for DelR and n Achievement were found. The inverse relationship between preference for DelR and acquiescence was, as predicted, also obtained. n Achievement was inversely related to acquiescence but only as a statistically nonsignificant trend. Responses to a simple open-ended aspiration measure, appropriate for use in other cultures, were found to significantly differentiate high and low n Achievement subjects as measured by the standard procedure, more occupational aspirations being

given by high n Achievement subjects. The possible usefulness of this aspiration measure as an alternative index of n Achievement in field work was suggested. Data on the interrelationships between a measure of social responsibility and the other variables were also presented. The over-all findings support the usefulness of relatively simple choice situations for eliciting preference for ImR as opposed to DelR in the attempt to elaborate the variables relevant to the ability to delay immediate gratification for the sake of later, larger outcomes, within the context of another culture. The present findings serve to elaborate the network of correlates of choice preferences for delayed, larger gratifications when these are opposed to smaller but immediately available rewards.

REFERENCES

ARONSON, E. The need for achievement as measured by graphic expression. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. Pp. 249–265.

ATKINSON, J. W. *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53,** 296–299.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60,** 151–174.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. Pp. 242–248.

FREUD, S. *Beyond the pleasure principle.* New York: Boni & Livewright, 1922.

HARRIS, D. B. A scale for measuring attitudes of social responsibility in children. *J. abnorm. soc. Psychol.*, 1957, **55,** 322–326.

LEVINE, M., GLASS, H., & MELTZOFF, J. The inhibition process, Rorschach human movement responses and intelligence. *J. consult. Psychol.*, 1957, **21,** 41–45.

LEVINE, M., SPIVACK, G., & WRIGHT, B. The inhibition process, Rorschach human movement responses and intelligence: Some further data. *J. consult. Psychol.*, 1959, **23,** 306–312.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MAHONE, C. H. Fear of failure and unrealistic vocational aspirations. *J. abnorm. soc. Psychol.*, 1960, **60,** 253–261.

MAHRER, A. R. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, **52,** 101–105.

MELIKIAN, L. Preference for delayed reinforcement: An experimental study among Palestinian Arab refugee children. *J. soc. Psychol.*, 1959, **50,** 81–86.

MESSICK, S., & JACKSON, D. N. Authoritarianism or acquiescence in Bass' data. *J. abnorm. soc. Psychol.*, 1957, **54,** 424–426.

METZNER, R. Preference for delayed reinforcement: Some complications. Harvard Psychological Clinic, 1960. (Mimeo)

MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *J. abnorm. soc. Psychol.*, 1958, **56,** 57–61.

MISCHEL, W. Preference for delayed reinforcement: Further cross-cultural applications. Harvard Psychological Clinic, 1959. (Mimeo)

MISCHEL, W. Preference for delayed reinforcement and social responsibility. *J. abnorm. soc. Psychol.*, 1961, **62,** 1–7.

RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

SAMELSON, F. The relation of achievement and affiliation motives to conforming behavior in two conditions of conflict with a majority. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. Pp. 421–433.

SINGER, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19,** 259–266.

SINGER, J. L., WILENSKY, H., & MCCRAVEN, VIVIAN G. Delaying capacity, fantasy and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, **20,** 375–383.

SPIVACK, G., LEVINE, M., & SPRIGLE, H. Intelligence test performance and the delay function of the ego. *J. consult. Psychol.*, 1959, **23,** 428–431.

(Received April 25, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 553–558

INTERFERENCE EFFECTS WITH SCHIZOPHRENIC SUBJECTS IN THE ACQUISITION AND RETENTION OF VERBAL MATERIAL

JOHN W. DONAHOE, MARY ELLEN CURTIN
University of Kentucky
AND LEONARD LIPTON[1]
Veterans Administration Hospital, Lexington, Kentucky

MEDNICK (1958) and Chapman (1958) have suggested that the aberrant language and conceptual behavior found in schizophrenia are, in part at least, due to the competition and intrusion of inappropriate associative responses. These investigators differ, however, in the assumptions underlying this assertion. Mednick, using Hullian learning theory (Hull, 1943) as his frame of reference, assumes that schizophrenia is characterized by heightened anxiety (D), which leads to an increase in the habit strength of remote associations within the response hierarchy. Thus, in simple tasks, i.e., those with few competing responses, schizophrenic performance should be superior to that of normals, whereas in more complex situations the reverse should hold. Mednick and DeVito (1958) support this contention. They required schizophrenics and normals to learn a list of paired-associates of low complexity (e.g., table-chair, dark-light, etc.) and one of high complexity (e.g., table-gas, comfort-chair, etc.). Complexity, therefore, was equated with the extent to which an S-R pair was in keeping with the cultural frequency of association between them (Russell & Jenkins, 1954). Schizophrenics learned the list of low complexity more quickly than did normals, whereas the rank order of groups was significantly reversed with the high complexity list.

For Chapman (1958), the interference hypothesis appears to have arisen chiefly from clinical observation (Arieti, 1955; Cameron, 1944) and, as such, is the result of empirical generalization rather than theoretical deduction. Chapman used a conceptual card-sorting task in which the subject was required to place an object-name (e.g., hat) in one of three categories: a member of the same class (cap), an associatively linked object (head), or an irrelevant object (street). Schizophrenics were found to commit more errors of the associated object type than were normals, especially with an increase in the difficulty of concept (Chapman, 1958). Ignoring the differences in rationale, the results of the Mednick and Chapman studies are quite compatible.

The generality of the interference hypothesis is open to question, however. The Chapman and Mednick studies have certain features in common: The S-R associations were learned outside of the experimental situation (probably prior to the onset of psychosis), and the findings therefore represent some degree of confounding of acquisition and retention. Also, the associations were meaningful, a form of association which many authors interpret as being dependent upon the mediation of implicit responses (Osgood, Suci, & Tannenbaum, 1957).

In an effort to clarify the range of valid application of the interference hypothesis, the present investigation explored the susceptibility of schizophrenics to interference effects in both the acquisition and retention of a set of rote verbal habits. In a serial learning situation, amount of interference in acquisition was manipulated by means of intralist similarity, while the analogous function in retention was served by varying the interlist similarity according to the retroaction paradigm. This approach has the further advantage of employing an experimental technique about which much is known.

METHOD

Subjects

Seventy-two subjects were used in the experiment, 36 normals and 36 schizophrenics. The normals were

[1] The authors wish to express their gratitude to the staff of the Veterans Administration Hospital, Lexington, for their assistance in permitting the study to be performed at that facility and for their assistance in recruiting subjects.

TABLE 1
EXPERIMENTAL LISTS

NI	M	NII
XUB	MAN	YOM
FOJ	DOOR	VEP
QIY	END	ZAW
KEJ	YEAR	JEG
YUN	GLASS	ZUX
SIH	PLAY	KUH

selected from students enrolled in laboratory sections of introductory psychology; the schizophrenics were drawn from a population of hospitalized veterans. No subject had previously participated in a verbal learning study. The schizophrenic group had a mean of 11.9 years of schooling[2] and a mean age of 31.8 years. All patients were receiving some form of tranquilizing medication at the time of testing, and all but 10 were from so-called chronic wards. The dosage for all but 11 of the patients was less than 100 milligrams per day. No attempt was made to differentiate among diagnostic subtypes, but case history material was examined to remove those who had ever carried a diagnosis other than schizophrenia.

Experimental Materials

As learning was expected to proceed at the slowest rate in the condition in which schizophrenics were required to learn nonsense syllables, an attempt was made to select materials for this condition first. A list of 13 nonsense syllables of 0–20% association value (Glaze, 1928) was composed and presented to six patients at a 3-second intersyllable rate. No subject could learn the list to a criterion of one perfect repetition. The original list was then progressively shortened until it had decreased to only 6 syllables (five anticipations) in length. Even with this drastic reduction, schizophrenic subjects found the task exceedingly difficult. Accordingly, a new strategy was adopted. In place of a machine-paced rate of presentation, subjects were given a switch which permitted them to expose the syllables at a self-determined rate. Under these conditions, a list of 6 syllables was learned in a reasonable number of trials.

Three sets of syllables were then constructed—a nonsense list (NI), a meaningful list (M), and a second nonsense list (NII) to be interpolated between the learning and relearning of NI. The meaningful words were monosyllabic and occurred with a frequency of at least 100 per million (Thorndike & Lorge, 1944). The items contained in each of the three lists are presented in Table 1.

Procedure

The schizophrenic and normal samples were each randomly divided into two groups, one of 24 subjects and one of 12 subjects. The groups of 24 subjects learned NI; the groups of 12 subjects learned M. This provided for a comparison of the relative performance of schizophrenics and normals as a function of intralist similarity. The groups which had learned NI were then further split into two subgroups of 12 subjects each. One of the subgroups received NII as the interpolated task, while the other subgroup received M. All who had learned NI were then required to relearn it. In every case, the acquisition criterion was one perfect repetition, and the interlist rest interval was 2 minutes. The recall and relearning data permitted a comparison between groups with respect to the retroactive interference effects of highly similar (NII) and highly dissimilar (M) interpolated material. Standard rote serial learning instructions were employed. The experimental session rarely exceeded 30 minutes, thus, diminishing the possibility of intrasession behavioral changes in the schizophrenics. Time per trial as well as errors and trials to criterion were recorded.

TABLE 2
ANALYSIS OF TRIAL, ERRORS/TRIAL, AND RATE MEASURES IN LEARNING TASKS

Source	*df*	Trials *F*	Errors/Trial *F*	Seconds/Trial *F*
Groups (G)	1	9.99**	14.84**	1.17
Lists (L)	1	31.66**	110.39**	28.71**
G × L	1	0.21	4.90*	0.05
Within cells	68	(9.58)	(0.30)	(235.62)
Total	71			

* $p < .05$.
** $p < .01$.

RESULTS[3]

Learning Tasks

Schizophrenics required significantly more trials than normals to attain the criterion of one perfect repetition (see Table 2, Trials column). However, contrary to what would be expected from the interference hypothesis, this difference was not more pronounced with NI than with M (G × L insignificant). For both lists, one-tailed t tests were significant between groups; t (46 df) = 2.64, $p < .01$ and t (22 df) = 1.87, $p < .05$ for NI and M, respectively. By the trials variable, therefore, group differences remained relatively constant in spite of a marked difference in list difficulty.

As the difference between groups in total error score was spuriously inflated due to

[2] With a sample of 315 schizophrenic patients selected from the same hospital, Richard M. Griffith has estimated the correlation between W-B IQ and years of schooling to be .62 ($p < .001$).

[3] The authors are indebted to the University of Kentucky Computing Center for making their facilities available.

TABLE 3

MEAN TRIAL, ERRORS/TRIAL, AND SECONDS/TRIAL MEASURES IN LEARNING, INTERPOLATION, AND RELEARNING

Groups	Measures	Learning		Interpolation		Relearning NI	
		NI	M	NII	M	NII	M
Schizophrenic	Trials	8.8	4.1	5.9	2.6	5.0	1.8
	Errors/Trial	2.6	1.4	2.3	0.9	1.9	0.3
	Seconds/Trial	55	31	47	32	46	34
Normal	Trials	6.2	2.2	5.1	1.8	3.5	1.4
	Errors/Trial	2.3	0.5	1.9	0.8	1.8	0.3
	Seconds/Trial	54	27	43	29	38	31

unequal opportunities to err resulting from an inequality in trials to criterion, the derived measure of errors per trial was used. An analysis of this measure revealed a higher rate of error production in the schizophrenic group (see Table 2, Errors/Trial column). Again, the schizophrenics achieved significantly higher scores on both NI, t (46 df) = 2.00, $p < .05$ and M, t (22 df) = 3.91, $p < .01$. An examination of the type of error as a function of groups resulted in an interesting finding. While no qualitative differences were detected in the errors of commission (intra- and extra-list intrusions), errors of omission (failures to respond) were much more frequent among schizophrenics. In fact, when this latter type of error was removed from the total error score, all error sources of variance fell to insignificance with the exception of lists, F (1, 68 df) = 29.73, $p < .001$. Thus, in the learning task, it would appear that group differences arose from a general decrement in the response tendencies of the schizophrenic group.

It will be recalled that with a machine-paced rate of syllable presentation, schizophrenics found serial learning to be an extremely difficult task. Accordingly, it could be argued that the schizophrenic has a slower self-presentation rate than the normal and that this reduced rate may compensate for the deleterious effects of response competition. This was not the case. The sole significant rate effect was associated with the differences between lists (see Table 2, Seconds/Trial column).

TABLE 4

ANALYSIS OF TRIAL, ERRORS/TRIAL, AND RATE MEASURES IN INTERPOLATION TASKS

Source	df	Trials F	Errors/Trial F	Seconds/Trial F
Groups (G)	1	2.57	1.70	0.40
Lists (L)	1	41.03**	39.00**	9.33**
G × L	1	0.00	0.44	0.02
Within cells	44	(3.25)	(0.50)	(255.18)
Total	47			

** $p < .01$.

In order to assess the possible effects of tranquilizing medication upon performance, all of the above criterion measures were reanalyzed with levels of drug as a variable. In no instance was medication or any of its interactions a significant source of variance.

Interpolated Tasks

With NII and M as the interpolated tasks, the performance of schizophrenics and normals was statistically indistinguishable (see Table 4) by all measures (see Table 3, Interpolation column). Meaningful and nonsense lists both gave rise to significant group differences in the initial learning task. The additional insignificance of group differences in errors of omission during interpolation tends to substantiate the interpretation that failures to respond accounted for the group differences during original learning.

An analysis of qualitative error differences (intrusions of NI into NII) and drug levels led to no statistically significant conclusions.

Retention Task

As interference effects are primarily confined to the initial trials of relearning (Underwood, 1945), the number of syllables incorrectly recalled on the first relearning trial provides an appropriate index for a test of the interference hypothesis as applied to retention. By this measure, NII interfered with the recall of NI to a significantly greater degree than did M (see Table 5, Errors column). Thus, the necessary precondition for the test was established. However, when the data were examined for group differences, none was found. The number of syllables correctly recalled was

TABLE 5
Analysis of Error and Time Measures in Recall and Trials Measure in Relearning

Source	*df*	Recall		Relearning Trials
		Errors *F*	Seconds/Trial *F*	*F*
Groups (G)	1	.66	.69	4.83*
Interpolated Lists (L)	1	75.61**	5.40*	40.83**
G × L	1	.12	.82	1.95
Within cells	44	(1.55)	(536.93)	(2.09)
Total	47			

* $p < .05$.
** $p < .01$.

essentially the same for both groups (2.6 for schizophrenics and 2.9 for normals); more importantly, the difference between groups was invariant over lists (G × L insignificant). An analysis of the rate of syllable presentation during the recall trial also revealed no group differences (see Table 5, Seconds/Trial column). The average duration of the recall trial was 49 seconds for schizophrenics and 43 seconds for normals. In short, neither the acquisition nor the recall data furnish support for the interference hypothesis.

When all the relearning data were appraised, the interpolation of NII was found to produce an increase in the number of errors per trial, F (1, 44 *df*) = 183.06, $p < .01$, and a decrease in the rate of syllable presentation, F (1, 44 *df*) = 5.90, $p < .05$, as compared to performance after the interpolation of M. Again, there were no group differences (see Table 3, Relearning column). An analysis of the trials data resulted in a somewhat different finding. In addition to the unequal relearning scores following the interpolation of NII and M, schizophrenics required significantly more trials to relearn NI than did normals (see Table 5, Trials column). The schizophrenic decrement was primarily due to performance after the interpolation of NII as shown by t tests between groups on each of the lists separately. Normals relearned NI following NII in fewer trials than schizophrenics, t (22 *df*) = 2.08, $p < .025$, whereas there was no significant difference between groups following the interpolation of M, t (22 *df*) = 0.80, $p > .05$. A comparison of performance between lists NII and NI provides supplementary information. While normals relearned NI in fewer trials than they learned NII, t (11 *df*) = 3.29, $p < .01$, schizophrenics required the same number of trials to relearn NI as to learn NII, t (11 *df*) = 1.29, $p > .05$. As there was no evidence of a group difference in recall, the relearning differences reflect a general performance decrement in the schizophrenics unrelated to interference effects. This decrement is most pronounced in the NI-NII-NI sequence, the condition in which the level of difficulty remained high throughout the experiment.

No significant group differences in errors of commission (intrusions of NII into NI) or omission were found in either recall or relearning. As was the case in the learning and interpolation tasks, no relationships were obtained between drug levels and any of the experimental variables.

Discussion

The interference hypothesis, as separately advanced by Mednick and Chapman, states that the behavior of schizophrenics is adversely affected to a disproportionate degree by increased response competition. The results reported here do not support this position. Although increased intralist and interlist similarity markedly interfered with acquisition and recall, respectively, the schizophrenic and normal groups were equally affected. The primary factor accounting for these discrepant findings appears to be the source of the response competition. Whereas Chapman and Mednick selected distracting material from response hierarchies formed through preexisting differential reinforcement, the stimulus material (nonsense syllables) employed here had not been previously discriminated. Competition, therefore, was not a product of meaningful similarity, but was largely engendered by primary stimulus similarity among the syllables.

Other evidence tends to support the conclusion that the interference hypothesis is tenable only with respect to competition arising from meaningful similarity. Ahana (1955) studied the acquisition of a serial list of adjectives by schizophrenics and normals. For half of each group, distracting nonsense

symbols were placed within a rectangular border which surrounded the items of the list. Acquisition was retarded by the distractors, but both groups displayed an equal decrement.

The assumption that interference phenomena, where they do occur in schizophrenia, result from heightened anxiety is also of questionable validity. O'Connor and Rawnsley (1959) found schizophrenics and normals to be no different in the acquisition of the conditioned eyeblink, whereas normals of high anxiety have been shown to condition more rapidly (Spence & Taylor, 1951). The same authors reported, however, that *verbal* conditioning was inferior in schizophrenics. In work concerned with primary stimulus generalization along the pitch continuum, Garmezy (1952) found no difference between schizophrenics and normals under nonstress conditions.

In the present experiment, those group differences which did emerge were apparently the conseqeunce of factors other than a differential susceptibility to interference effects. In the initial learning task, the performance of schizophrenics was inferior to that of normals with both the meaningful and nonsense materials. A closer examination of the data revealed that the difference was largely due to a general failure of the schizophrenics to respond. In the relearning task, the schizophrenics were again inferior to the normals. This finding was not a function of differences in recall, however, as the errors committed on the first trial of relearning failed to discriminate between the groups. In this respect, the data are similar to those reported in an early study by Hull (1917). The relearning differences were primarily confined to the experimental sequence of greatest difficulty and are perhaps analogous to fatigue effects. This supposition is strengthened by the work of Bleke (1955), who reported greater reminiscence effects in schizophrenics under conditions in which wrong responses led to verbal punishment. Shakow (1946) has also found reminiscence effects in schizophrenics on the pursuit rotor. It should be noted that both of these factors, the decreased readiness to respond and the brevity of performance, make for difficulties in attempting to determine whether schizophrenia is characterized by a basic conceptual deficit.

Another variable influencing the behavior of the schizophrenic appears to be the manner in which the experimental material is presented. Although not studied systematically, preliminary work strongly suggests that performance deteriorates inordinately if the schizophrenic is not allowed to program the stimuli himself. A psychophysical study of brightness discrimination by Wohl (1958) leads to the same conclusion. Using the method of average error, the difference limen was the same for both schizophrenics and normals. With the method of constant stimuli, the difference limen for the schizophrenics was larger. The experimental findings probably reflect the waxing and waning of attention so frequently reported by clinical observers as a predominant feature of schizophrenia.

SUMMARY

Two groups of subjects, 36 normals and 36 schizophrenics, were studied in the acquisition and retention of serial lists of varying intralist and interlist similarity. An examination of the data led to the following conclusions:

1. The heightened sensitivity of schizophrenics to interference and distraction is confined to response competition based upon meaningful similarity.

2. Prolonged warm-up and brevity of actual performance make for grave difficulties in achieving a definitive statement concerning a basic conceptual deficit in schizophrenia.

3. Experimental situations in which the subject controls the time of presentation of the critical stimuli are of particular benefit to schizophrenics.

REFERENCES

AHANA, ELLEN. Verbal rote learning in schizophrenic and normal subjects. Unpublished doctoral dissertation, Northwestern University, 1955.

ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.

BLEKE, R. C. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, **23**, 479–498.

CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. II. New York: Ronald, 1944. Pp. 861–921.

CHAPMAN, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 374–379.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **21**, 253–276.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 26–35.

HULL, C. L. The formation and retention of associations among the insane. *Amer. J. Psychol.*, 1917, **28**, 419–435.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, **55**, 316–327.

MEDNICK, S. A., & DEVITO, R. Associative competition and verbal learning in schizophrenia. Paper read at Eastern Psychological Association, Philadelphia, April 1958.

O'CONNOR, N., & RAWNSLEY, K. Two types of conditioning in psychotics and normals. *J. abnorm. soc. Psychol.*, 1959, **58**, 157–161.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. NS onr-66216)

SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. dis. Monogr.*, 1946, No. 70.

SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Bureau of Publications, Teachers College, 1944.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 73).

WOHL, S. A. An investigation of perceptual sensitivity and psychophysiological characteristics of attention. Unpublished doctoral dissertation, University of Kentucky, 1958.

(Received May 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 559-564

DISAGREEMENT WITH OTHERS, THEIR CREDIBILITY, AND EXPERIENCED STRESS

HAROLD B. GERARD
Bell Telephone Laboratories, Murray Hill, New Jersey

To the extent that a person (ego) is not able to account for some judgmental disagreement with another (alter), he experiences stress. A large body of literature has examined various methods by which ego may reduce this stress through resolving the disagreement with alter (e.g., Back, 1951; Festinger 1950). Among these methods is for ego to employ information he may have regarding his ability relative to alter. A number of studies have specifically manipulated this factor (Festinger, Gerard, Hymovitch, Kelley, & Raven, 1952; Festinger & Thibaut, 1951; Gerard, 1953) and have shown a tendency on the part of ego to reject extreme opinions when he believes that his ability might possibly differ from another holding an opinion very discrepant from his own. Gross trends toward group subdivision were discovered in this earlier research where ego presumed the existence of an ability differential. The present study shifts the focus by inquiring into processes occurring at the individual level that may account for this tendency observed in groups.

When ego is confronted by disagreement, the correctness of his judgment is brought into question and he must decide how much credibility to attach to his judgment and to alter's. If he has evidence indicating that his ability is superior to that of alter, he should attach little credibility to alter's judgment; the impact of the disagreement is thus reduced for him. Similarly if ego believes himself to be inferior to alter, a basis for resolving the disagreement also exists; that is, he may attach little credibility to his own judgment. If, on the other hand, he were provided with a basis for attributing equal credibility to his own and to alter's judgment he should find it relatively difficult to resolve the discrepancy.

When confronted by disagreement we have said that ego should tend to compare his ability with that of alter. He should also expect alter to be making a similar comparison. Given a basis for an ability differential, the evaluation ego has of himself, the evaluation he has of alter, and the evaluation he assumes alter to have of him should tend to become internally consistent. The criterion of consistency here is the balance criterion of Heider (1958) and Cartwright and Harary (1956). We have, however, made our performance evaluation psychologically interactive by endowing them with behavioral implications. This dynamic involves an assumption for which there is considerable evidence: that ego's disposition toward alter conditions his acceptance or rejection of alter's opinions and evaluations, including the evaluation he assumes that alter has of him. Inconsistency would obtain, for example, if ego evaluated both himself and alter positively and he assumed that alter evaluated him negatively. Under these circumstances his evaluation of himself would tend to be equivocal. Being positively disposed toward alter would, according to our assumption, induce ego to accept the evaluation he believes that alter has of him, one that conflicts with his own high self-evaluation. A strain toward consistency or balance should exist that should tend to produce a change in one or more of the evaluations. After having been provided with a basis for explaining the discrepancy we would expect ego to develop and maintain an unequivocal self-evaluation, which, according to our argument, should be manifested in internally consistent self-other evaluations.

A previous experiment (Gerard & Rotter, 1961) found that internally consistent evaluations were held by subjects who yielded a great deal and by those who maintained their own judgment in the face of disagreement. No attempt was made to vary the subject's relative ability. The present suggestion, however, is that accepting or rejecting the others' judgments reflects a situation in which the subject attributes an ability difference between himself and the others whereas a lack of behavioral resolution, reflected by yielding only part of the time, reflects a situation in which the subject

believes himself to be about equal to the others in ability.

Previous research (Festinger et al., 1952; Festinger & Thibaut, 1951; Gerard, 1953) has shown that as judgmental discrepancy increases so do tendencies toward resolving the discrepancy. Resolution should therefore be more rapid, and be accompanied both by a reduction of stress and an increase in evaluation consistency.

The foregoing discussion has advanced the following hypotheses:

1. When confronted by judgmental disagreement with alter, ego should experience less stress where an ability differential exists as compared to a condition where alter is presumed to be equal in ability.

2. When confronted by disagreement with alter, ego should develop and maintain greater self-other evaluation consistency where an ability differential exists as compared to a condition where alter is presumed to be equal in ability.

3. The greater the discrepancy confronting ego, the more rapidly should the tendencies described in Hypotheses 1 and 2 become manifest.

METHOD

Subjects

Eighty paid volunteers from Drew University served as subjects. Each experimental group was either all male or all female, the sex factor being balanced across the design.

Procedure

The stimuli developed by Asch (1951) were used. Each consists of a single line presented at the left of the subject's visual field and three comparison lines of differing length presented at the right, one of which is equal in length to the single line. The subject's task is to determine which of the comparison lines is equal in length to the single line. The differences among the comparison lines are rather gross, making the task easy. Two series of nine stimuli each were projected successively on a screen 20 feet in front of the subject. For each stimulus the subject depressed one of three keys depending upon which comparison line he chose. His judgments and what he presumed to be the judgments of three other subjects registered on a panel of lamps in front of him. In actuality the others' judgments were preprogramed by the experimenter such that on Trials 1, 2, and 5 their judgments were correct whereas on the other six trials in each series they were unanimous in their choice of one of the two incorrect comparison lines.

The four subjects in each group were seated in adjoining cubicles and each was privately informed that he would be identified as subject Number 3. He did not know the personal identity of the others whom he believed had been given the identification Numbers 1, 2, and 4.

The sequence of events was as follows: A slide was projected. After 8 seconds, the subject was asked to make his choice and to keep his choice key depressed. The display was then turned on. The "others' judgments" were identically displayed in all four cubicles. After 4 seconds the display was turned off and the subject was asked to release the key and await the next slide. This procedure served to confront each subject with the fact that he was a deviate in a group of four. The simultaneous presentation of the judgments precluded the possibility of yielding.

After each series of judgments the subject answered a questionnaire designed to determine the degree of consistency of his self-other evaluations.

Experimental Variations

In order to induce an ability differential the subject was given another judgmental task prior to the line judgments. The stimuli for this series were visual mazes developed by Jackson and Salzstein (1958). All the mazes consisted of identical circular figures in which the same four curved lines, originating from a common point on the circumference, proceeded across the interior of the circle to four other points on the circumference. The subject's task was to choose the shortest of the four possible paths through the interior of the maze from the start—the common beginning of all four paths—to an arrowhead variously located on the circumference. Part of the circumference, between the end of the interior path and the arrowhead, was to be included in the estimate of length. The only difference among stimuli was in the placement of the arrowhead. We pretested 96 of these mazes on 20 subjects not included in our experimental sample and selected 10 that were judged to be difficult by most subjects. After judging these 10 mazes the subject was scored as to how many correct choices he had made. He was told that most people like to know how they compare with others in this ability. We were therefore returning his answer sheet with his score and the scores of the three others. Ten points were given for each correct choice so that a perfect score was 100. In order to manipulate differential ability, false scores were given to the subject. In all cases he himself received a score of 70 but the scores of the others were varied in order to create the impression of differential ability. His own score was encircled in red. No indication was given as to either the name or identification number of those making the other scores. In one treatment the others scored 30, 40, and 50 (Others-less), in a second treatment they were given as 60, 70, and 80 (Others-equal), and in the third they were 90, 100, and 100 (Others-greater). The subject was told that the same ability is involved in making the judgments that were to follow.

In order to examine the hypothesized effects of an increase in resolution tendencies, we varied the extent to which the others disagreed with reality on the critical

TABLE 1
PERFORMANCE EVALUATIONS FOR BOTH SERIES OF JUDGMENTS

Ability of Others Relative to Self	Evaluation of Self				Evaluation of Others				Other's Evaluation of Self			
	Small Discrepancy		Large Discrepancy		Small Discrepancy		Large Discrepancy		Small Discrepancy		Large Discrepancy	
	1st Series	2nd Series	1st Series	2nd Series	1st Series	2nd Series	1st Series	2nd Series	1st Series	2nd Series	1st Series	2nd Series
Less	2.53	2.33	2.20	2.28	2.47	2.53	3.20	3.00	3.07	3.53	3.64	3.58
Equal	2.67	2.67	2.54	2.90	2.50	2.50	2.62	2.40	3.17	3.42	3.31	3.60
Greater	3.92	3.75	3.13	2.80	2.31	1.92	2.00	2.47	3.85	3.62	3.78	3.78

trials. In the Large Discrepancy treatment the others appeared to choose the more incorrect of the two incorrect comparison lines whereas in the Small Discrepancy treatment they chose the one which was closer in length to the correct line.

Measurement of Stress

With the aid of a switching arrangement we recorded the base level skin resistance of our four subjects on a single recorder, the C. H. Stoelting psychogalvanograph, by rotating the one instrument to sample each subject periodically. The subject was attached across the input with Scherping type finger electrodes to which a paste of glycerine, bentonite, and salt water had been applied. The rationale given the subject for using these electrodes was that we were interested in examining his physiological reactions while making the judgments.

RESULTS

Experimental Manipulations

On the questionnaire administered after each series of judgments the subject evaluated his own ability and the ability of the others, and guessed how the others would evaluate him. The same six-point scale, which ranged from a very positive to a very negative evaluation, was used for each of the three ability estimations. These data are presented in Table 1. Recall that the ability manipulation varied the scores of the others while the subject's own score was fixed at 70. Under Small Discrepancy we find, for both series, little difference in the ability attributed to the others. The differential emerges in self-evaluation ($p < .01$ by an F test). This does not obtain under Large Discrepancy where we see a linear effect of ability upon the evaluation of the others for the first series ($p < .05$), an effect that disappears, however, for the second series. We also note that under Large Discrepancy the linear effect upon self-evaluation, which was significant for the first series ($p < .01$), tends to be reduced. No consistent effects appear for the other's evaluation of the self.

There is only indirect evidence on the success of the discrepancy manipulation. On each questionnaire the subject was asked to indicate the number of judgments concerning which there was disagreement. The average perceived disagreement under Small Discrepancy was 4.59 trials as compared to 5.78 for the Large Discrepancy treatment ($p < .001$ by t). These figures represent the combined average for each of the two series. The extent of the disagreement on the critical trials thus appears to have been reflected in the perceived frequency of disagreement.

Stress

In view of metric considerations discussed by Lacey and Siegal (1949) we converted our resistance (R) measurements into conductance (C) as follows:

$$C(\mu\text{mhos}) = \frac{1}{R(\text{ohms})} \times 10^6$$

Table 2 presents the change in base conductance from the end of the maze judgments to the end of the first series of line judgments and then to the end of the second series of line judgments. There were no significant differences among treatments at the end of the maze judgments. In spite of a polarizing tendency associated with the type of electrodes used, conductance increases over time ($p <$ .001 by a sign test for each series). The rate of increase, however, decreases from the first to the second series ($p < .01$ by t). The ability variable shows an effect in that for the first series the increase in conductance is greater the *lower* the ability of the others relative to the

TABLE 2
CHANGE IN BASE LEVEL OF SKIN CONDUCTANCE IN MICROMHOS

Ability of Others Relative to Self	Small Discrepancy		Large Discrepancy	
	1st Series Change	2nd Series Change	1st Series Change	2nd Series Change
Less	+3.55	+0.48	+3.00	+2.47
Equal	+2.73	+1.62	+1.67	+1.44
Greater	+1.97	+1.41	+1.38	+1.47

TABLE 3
PROPORTION OF SUBJECTS HAVING BALANCED SELF-OTHER EVALUATIONS

Ability of Others Relative to Self	Small Discrepancy			Large Discrepancy		
	1st Series	2nd Series	d	1st Series	2nd series	d
Less	.73	.73	.00	.46	.50	+.04
Equal	.50	.67	+.17	.77	.90	+.13
Greater	.69	.69	.00	.93	.93	.00

subject (the linear component is significant at beyond the .01 level by an F test). This trend continues for the Large Discrepancy treatment ($p < .01$) whereas under Small Discrepancy we note a leveling off tendency where the others are lower in ability (the increase is only 0.48). The increase for the first series tends to be somewhat greater under Small as compared to Large Discrepancy ($p < .10$).

Consistency of Self-Other Evaluation

We see from Table 1 that the subjects tended to select their evaluation alternatives from the positive end of the six-point performance scale, the over-all median occurring between the second and third scale points. We therefore labeled an evaluation of 1 or 2 as positive and 3 through 6 as negative, evaluations of 5 and 6 being very infrequent. The Cartwright and Harary (1956) criterion applied here would classify as balanced those instances in which all three evaluations are positive along with those in which one is positive and the other two are negative. Of the eight possible configurations (each of the three evaluations can be either positive or negative), four are balanced. The balanced evaluations are the consistent ones, that is, those in which an unequivocal self-evaluation obtains. Table 3 shows the proportion of balanced evaluations occurring in each experimental condition for both series of judgments. Under Small Discrepancy the proportion for the equal ability treatment is somewhat lower than for either of the other treatments ($p < .20$ by chi square). This difference disappears after the second series. Under Large Discrepancy a linear tendency obtains. Where the others are lower in ability the relatively small proportion of balanced evaluations does not change after the second series. The other two treatments become equated. We also note that where there tends to be a greater proportion of balanced evaluations, the proportion tends to be somewhat greater in the Large as compared to the Small Discrepancy treatment.

The foregoing analysis can be refined if we assume that where the subject made very positive or very negative evaluations the degree of consistency would be felt more strongly, so to speak, than where his evaluations were more moderate. With the aid of this assumption, we derived the following measure of the degree of balance: an evaluation of 1 was scored +2 and 2 was scored +1, reflecting the fact that the former was a more positive evaluation. An evaluation of 3 was scored −1 and 4 through 6 were scored −2. Balance scores of +8, +4, +2, +1, −1, −2, −4, and −8 were therefore possible. These scores reflect, we propose, the degree to which the evaluations are consistent. Using this measure we determined the mean degree of consistency in all conditions for both series (Table 4). The trends here are generally consistent with those in Table 3. Under Small Discrepancy we note a curvilinear trend with ability ($p < .10$ by t) for the first series which appears to accentuate somewhat for the second series ($p < .05$ by t). Under Large Discrepancy a relatively high degree of consistency appears in the Others-

TABLE 4
MEAN DEGREE OF OVER-ALL CONSISTENCY AMONG SUBJECT'S EVALUATIONS

Ability of Others Relative to Self	Small Discrepancy			Large Discrepancy		
	1st Series	2nd Series	d	1st Series	2nd Series	d
Less	2.27	2.00	−0.27	0.27	0.17	−0.10
Equal	0.92	1.06	+0.14	2.85	3.80	+0.95
Greater	1.15	2.42	+1.27	2.70	4.15	+1.45

equal and Others-greater treatments as compared with the Others-less treatment ($p < .01$ by t). This difference, too, appears accentuated over time ($p < .001$ by t). A trend which appears to be consistent for both discrepancy treatments is the linear effect of ability upon change in the degree of balance ($p < .01$ by F).

DISCUSSION

Hypotheses 1 and 2 appear to be substantiated only under Small Discrepancy where curvilinear trends were noted for both the conductance and consistency data. During the first series we saw a linear effect of ability upon conductance for both discrepancy levels. Under Small Discrepancy there was evidence of a leveling off during the second series in both ability treatments where a differential existed. In these two treatments greater consistency of evaluation was also observed, although in the Others-less treatment a continued increase in consistency was not in evidence. A fact that is possibly important in explaining some incongruence between the conductance and consistency data concerns the timing of these measures. Conductance change is based upon skin resistance readings taken during the actual confrontation whereas the consistency data were obtained from a questionnaire administered after the subject had experienced the disagreement. Between the end of each series of judgments and the time the subject actually made his questionnaire responses, time had elapsed during which he had a chance to reflect upon what had happened and bring his evaluations into line with events. Thus in the Small Discrepancy–Others-less condition, the lack of coincidence between the conductance and consistency data might be accounted for by this time lapse.

Under Large Discrepancy, which shows the same linear effect of ability upon conductance change for the first series, the same leveling off trend does not appear in the Others-less treatment. Self-evaluation consistency likewise shows a linear effect for the first series that is sustained over time. Perhaps the most we can say here is to suggest that partial support for Hypotheses 1 and 2 exists under Large Discrepancy in the comparison between the Others-greater and Others-equal conditions.

We note some support for Hypothesis 3, which predicts both greater relative stress reduction and greater evaluation consistency under Large Discrepancy. The evidence shows less increase in stress under Large Discrepancy and also, where relatively high consistency does obtain (under the Others-equal and Others-greater treatments), the level is higher than under Small Discrepancy. Again we are faced with the results for the Large Discrepancy–Others-less condition which do not coincide with the prediction.

In attempting an ex post facto interpretation of the discrepant results, two facts assume crucial importance. The first is that even though the ability manipulation varied the others' scores and held the subject's score fixed, the subject did the reverse. He apparently used their scores as a benchmark or average against which to judge his own performance. His own score was labile relative to the others. We thus have a situation in which the subject sees the others as of average ability and himself as less, more, or equally able.

The second fact, which requires underscoring, is that the subject found himself in disagreement with three independent judges. No social influence among the others was possible since all judgments were displayed simultaneously. This undoubtedly had considerable impact upon the subject and would tend to reduce the credibility of his own judgment.

What are the implications of these circumstances? We can assume that the degree to which the others' judgments were compelling was approximately the same for the three ability levels. The variable factor was the amount of credibility the subject could attach to his own choice. Seen in this light, the results may reflect increasing severity of a postdecisional predicament in which the subject must reconcile the disagreement with knowledge he has about himself. The more difficult it is for him to lower his commitment to the choice he had made, the greater the stress he is likely to experience and the more likely should his self-evaluation remain in question. These highly speculative remarks may help to understand the effects observed under Large Discrepancy. Under Small Discrepancy the subject may have been able to discount the others

relative to himself since there was greater room for doubt as to who was correct. Some weak support for this supposition appears if, in Table 1, we examine changes in evaluations from the first series to the second series. A slight, but statistically insignificant trend exists for the subject under Small Discrepancy to increase his self-evaluation *relative* to the others over time whereas the reverse obtains under Large Discrepancy.

In line with these remarks we find in Table 4 a linear trend in both discrepancy treatments of the effect of ability upon change in degree of consistency, a finding that would suggest that the same predicament occurred to some extent with increasing self-evaluation in both treatments.

This attempt at interpreting the unpredicted findings must be taken as suggestive. Research is in progress that studies, among other things, the effects of making the others' judgments truly labile, of varying the impact of the others' judgment, and of varying the time at which consistency measures are administered. In this research continuous skin resistance records are being made for each subject.

SUMMARY

In a setting where the subject was confronted with disagreement from three other subjects, we varied both the degree to which he was in disagreement with them and his relative ability to make the visual judgments in question. He was confronted with two series of nine judgments each and found himself in disagreement with the others on six of the nine trials. We determined, after each series, the internal consistency of his self-evaluation, his evaluation of the others, and the evaluation he guessed they had made of him. We also measured the stress the subject experienced as reflected by changes in his skin conductance. Where there was a clear differential in ability, the subject evidenced both a relative diminution of stress and greater relative consistency of self-other evaluation. This relationship held where he was in mild disagreement with the others. Where the disagreement was large, the greater his ability relative to the others, the more stress did he experience and the less consistent were his self-other attitudes. An interpretation for this finding is offered.

REFERENCES

ASCH, S. E. Effects of group pressures upon the modification and distortion of judgments. In M. H. Geutzkow (Ed.), *Groups, leadership, and men.* Pittsburgh: Carnegie Press, 1951. Pp. 117–190.

BACK, K. The exertion of social influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46,** 9–23.

CARTWRIGHT, D., & HARARY, F. Structural balance: A generalization of Heider's theory. *Psychol. Rev.*, 1956, **63,** 277–293.

FESTINGER, L. Informed social communication. *Psychol. Rev.*, 1950, **57,** 271–282.

FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. H. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5,** 327–346.

FESTINGER, L., & THIBAUT, J. W. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46,** 92–100.

GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1953, **6,** 249–271.

GERARD, H. B., & ROTTER, G. S. Time perspective, consistency of attitude, and social influence. *J. abnorm. soc. Psychol.*, 1961, **62,** 565–572.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JACKSON, J. M., & SALZSTEIN, H. D. The effect of person-group relationships on conformity processes. *J. abnorm. soc. Psychol.*, 1958, **57,** 17–24.

LACEY, O. L., & SIEGAL, P. S. An analysis of the unit of measurement of the galvanic skin response. *J. exp. Psychol.*, 1949, **39,** 122–127.

(Received May 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 565-572

TIME PERSPECTIVE, CONSISTENCY OF ATTITUDE, AND SOCIAL INFLUENCE

HAROLD B. GERARD AND GEORGE S. ROTTER
Bell Telephone Laboratories, Murray Hill, New Jersey

CONSIDERABLE attention has been devoted in recent years to the study of conformity. In the typical experiment the subject is confronted with a conflict induced by finding that his judgment concerning some object or event is different from the judgments of one or more other subjects. Interest has focused upon factors that tend to produce either independence or yielding. Subjects in these experiments are generally strangers to one another and assume that their association will terminate at the end of the experimental session. What would be the effect of projecting their association with each other into the future, beyond the one experimental session? What is the nature of the expectations that a person holds in regard to the other(s) when he anticipates future association?

The other may serve the person as a referent against which he may "check" his judgments and opinions. To the extent that he attributes to the other the skills or capacities necessary to make a judgment or to hold an opinion, the individual may rely upon him as a source of information about the object or event under consideration.

The other may also serve to mediate other satisfactions for him. In the case of institutionalized associations such as marriage, employment, schooling, and club membership, some of the satisfactions may be easily identified. Often, however, the bases for interdependence defy easy and immediate discovery. In casual human intercourse, the participants are continually appealing to each other's needs. The continuation of a given relationship will depend upon the extent to which the needs of the persons concerned are being met as compared with their respective estimates as to the possibility for meeting these needs via other activities. We have each learned that through association with others we do realize satisfactions of one kind or another. Thus, to the extent that a person projects an association with another into the future, he should, in general, anticipate dependence upon him. Disagreement between two individuals concerning something that each regards as important can be a threat to their continuing relationship. Indeed, there is a tendency over time for a person to ignore another whose opinion is very different from his own (Festinger, Gerard, Hymovitch, Kelley, & Raven, 1952; Schachter, 1955). In the face of disagreement we can expect a person to experience a greater force to resolve the disagreement, the longer the projected future association between them.

The nature of the projected future association should affect the individual's present disposition toward the other. If he expects their future association to be concerned with subject matter that is the same or similar to the subject matter with which they are currently concerned, he should estimate that the other's present and future judgments should be highly correlated. If their present judgments disagree and he expects the same activity to continue, he should anticipate future disagreement as well. In general, to the extent that the future holds similar subject matter in store, he should anticipate that present disagreement presages future disagreement. Given that he does not reject the other as a source of information, the more continuous the future (as regards subject matter) the more will the individual try to resolve any discrepancy of judgment that may arise.

In summary, the longer a person's projected future with another and the more similar he expects the future activities to be to present ones, the more readily should he attempt to adjust to present discrepancies of judgment with the other.

In addition to the effects of future perspective, the present experiment was also concerned with studying the relationship between the subject's self-other evaluations and his tendency to yield to the others' judgments. Let us assume that his evaluation of the other(s) conditions the subject's acceptance or rejection

of their opinions and evaluations. We can, with the aid of this assumption, designate as consistent or inconsistent the set of evaluations that includes the subject's evaluation of himself, his evaluation of the other, and the evaluation he believes the other has of him. To be consistent his self-evaluation must coincide with the evaluation he sees reflected by the other, this reflected evaluation being composed of two components. One component is the subject's evaluation of the other, that should condition the subject's acceptance of the other's evaluation of him, and the evaluation he believes the other has of him (cf. Gerard, 1961, for a more detailed discussion of direct and reflected self-evaluation). The Heider (1958) criterion of balance applies here. Of the eight possible configurations (each of the three evaluations can be either positive or negative), four are balanced or consistent and four are unbalanced or inconsistent. We would expect that a person who yields consistently or maintains his independence consistently should tend to be consistent in his self-evaluation. If a sequence of similar decisions (to yield or not to yield) is made, the process should be circular and cumulative. That is, consistency of evaluation should condition the decision which, in turn, should tend to produce greater consistency of evaluation.

METHOD

Subjects

Ninety-six Rutgers University undergraduates were used in the experiment as paid volunteers. Both men and women took part, the design being balanced for sex. In any given experimental group only men or only women were run.

General Setting

The present experiment employed a modified version of the so-called Asch (1951) situation. Subjects were tested four at a time and were seated in adjoining cubicles visually isolating them from each other. A slide consisting of a single line on the left and three comparison lines on the right, each of a different length, was projected on a screen 20 feet away. The subject's task was to choose from among the three comparison lines the one most nearly equal in length to the line on the left. One of the three lines was actually equal to the single line. On the subject's desk was a small envelope that contained a card with the subject's identification number on it. The experimenter asked the subject to look at the card and informed him that the numbers 1 through 4 were assigned to the cubicles at random in order to maintain personal anonymity during this phase of the experiment. After a stimulus was exposed the subject was asked to respond in sequence when his number was called. His response was made by pressing one of three keys corresponding to his choice of comparison line. The subject's choices and the choices of the others appeared, as they were made, on a panel of red lights in front of him. The panel consisted of a matrix of four rows (one for each subject) and 3 columns (one for each comparison line). The same series of nine slides was shown four times in succession. After each series there was a pause during which a brief questionnaire was administered.

Each subject was led to believe th. ' he and he alone was Subject #3 and he assumed that the responses he saw on his panel had been made by the other three subjects. Actually the responses of the others (1, 2, and 4) were fictitious and appeared in a preprogramed sequence such that "the others" were always in unanimous agreement and in disagreement with the correct comparison line on six of the nine trials. They "chose" the correct line on Trials 1, 2, and 5.

The subject received his instructions over a headset. The entire sequence of events during a series was electronically controlled by tones which were recorded on tape along with the verbal instructions. These tones, which were not heard by the subject, were used to operate the slide projector and to display each false judgment at the proper time. The subject's choices were automatically recorded in the form of tones on a second tape recorder.

Experimental Variations

When the subject signed up for the experiment, he volunteered to take part in four sessions on four different days. This four-session perspective was used as a basis for inducing different expectations about the future.

In order to induce a perspective of continued association with the others who were present, half of the groups were told that the same individuals would be together during all four sessions. The other half of the groups were told that the group composition would undoubtedly change from session to session.

The expectation of continuity of activity during all four sessions was induced by informing the group that we were concerned only with visual discrimination, and each of the sessions would therefore require similar judgments. The first session would concern estimating the length of lines, the second the size of areas within geometric figures, the third estimation of the length of curved lines, and the fourth estimates of the length of figures that had been distorted slightly. The experimenter then presented illustrations of all four types of judgment. To induce expectations of varied activities, the group was told that we were interested in various kinds of discrimination. The first session would concern comparing the length of lines, the second the quality of odors, the third lifted weights, and the fourth the length of musical passages. Illustrations of the first three types of judgment were presented.

The assignment of subjects was balanced in a 2 × 2 factorial design with 24 subjects in each of the four treatments: continued membership-continued task (MT), continued membership-discontinued task ($M\bar{T}$), discontinued membership-continued task ($\bar{M}T$), and discontinued membership-discontinued task ($\bar{M}\bar{T}$).

RESULTS AND DISCUSSION

We examine first the number of errors made by the subject in each treatment as an indication of a willingness to adjust to the discrepancy in favor of the others, then turn our attention to the attitudinal data.

Analysis of Conforming Errors

We predicted that the anticipation of continued future membership would engender dependency and, in the face of disagreement, greater willingness to yield to others; the tendency should be heightened with the anticipation of a continuous task. Table 1 presents the average number of yielding responses on each of the four series. An examination of the data for the effects of membership and task continuity does not present a simple picture. Over all 36 trials we observe a slight, statistically insignificant trend opposite to the predicted effect of membership perspective, (M vs. $\bar{M}$), whereas the effect of task perspective (T vs. $\bar{T}$) shows no trend.

Interesting details emerge, however, when the tendency to yield over time is examined. The M treatment showed little change over time, albeit minor fluctuations did occur from series to series. The $\bar{M}$ treatment presents quite a different picture. An initially high level of yielding was maintained for two series, but began to drop off during the third series, and dropped sharply at the end ($p < .01$ by t for the difference between Series I and IV). A plausible interpretation of these results is based on the assumption that the short run perspective produces a sense of urgency. Since there is only a limited time during which the sharp discrepancy in viewpoint can be resolved, the subject displays an all-out initial effort to compromise. When this fails after repeated tries, he comes to realize that the others will continue to disagree. The subject therefore begins to revert to his initial private judgments. The very sharp drop after the third series suggests that if the subject anticipated a fifth and a sixth series, his tendency to yield would probably continue to decrease rapidly. For the M treatment, however, we would expect little change.

TABLE 1

JUDGMENTAL ERRORS AS AFFECTED BY TASK AND MEMBERSHIP CONTINUITY AND SEATING ARRANGEMENT

Variable[a]	Series				
	I	II	III	IV	Total
M	1.69	1.52	1.60	1.56	6.37
$\bar{M}$	1.98	1.98	1.79	1.44	7.19
T	1.92	1.83	1.69	1.46	6.90
$\bar{T}$	1.75	1.67	1.71	1.54	6.67
MT	1.83	1.75	1.67	1.63	6.88
$\bar{M}T$	2.00	1.92	1.71	1.29	6.92
$M\bar{T}$	1.54	1.29	1.54	1.50	5.87
$\bar{M}\bar{T}$	1.96	2.04	1.88	1.58	7.46
Seat 1	1.33	1.42	1.08	1.08	4.91
Seat 2	1.42	1.83	1.67	1.13	6.05
Seat 3	2.04	2.04	1.88	1.79	7.75
Seat 4	2.54	1.71	2.17	2.00	8.42

[a] M = Continued membership, $\bar{M}$ = Discontinued membership; T = Continued task, $\bar{T}$ = Discontinued task.

A similar trend appears as a function of task continuity. However, the relationship is reversed. The T treatment showed a somewhat higher initial level than the $\bar{T}$ treatment, and then dropped off at an increasing rate from series to series. The $\bar{T}$ treatment, on the other hand, fluctuated from series to series with some tendency toward an over-all decrease. The initially higher level for the T treatment conforms to our expectation. The sharp decrease, however, may reflect the subject's reaction to an unbending majority ($p < .01$ by t for the difference between Series I and IV). An initial willingness to compromise with no evidence of reciprocation may have resulted in a strong tendency to reject the information proffered by the others. Also, given that tasks similar to the present one are anticipated in the future, it is important to establish initially a stable frame of reference so that the future tasks may be performed adequately. Since disagreement with the others was so marked, the subject tended to rely upon his

own visual sense impressions rather than those of the others. Had the disagreement been less pronounced, reliance upon the others rather than upon the self might have been enhanced, given a continuing task perspective.

We observe the sharpest decrease in errors over time in the $\overline{M}T$ treatment ($p < .01$ by F for the linear trend), which results, we have suggested, from the initial urgency of a short-term time perspective and subsequent rejection after unsuccessful overtures combined with the need to establish a stable visual frame of reference since the future presages similar judgments. Comparison with the MT treatment suggests the operation of constraints against rejection since the individual probably anticipates at least partial reliance upon the others as future sources of information. In comparing the $\overline{MT}$ and $M\overline{T}$ treatments, we again observe the differential decrease due to the membership variable. Both trends, however, are tempered by the effects of the anticipation of discontinuing the activity.

These results were, for the most part, unanticipated and the foregoing discussion must therefore be regarded as tenuous speculation as to the processes which underlie them.

An unanticipated and puzzling effect of the seating arrangement also occurred. A progressive increase in yielding occurs with seat number ($p < .01$ by F for the linear trend). There is a strong difference over all 36 trials and also over time from series to series, with the exception of Series II. Since there was a free choice of seat as the subjects entered the room, there may have been a personality predisposition in the choice of seat. The physical arrangement of the cubicles may have, in and of itself, produced the difference irrespective of possible self-selection. Cubicle 1 was closest to the door. The person in that cubicle, therefore, may have felt that he had easy access to an escape route so to speak. The person in Cubicle 4, on the other hand, had all three of the others between himself and possible escape. The greater the number of people between the subject and the door, the more "hemmed-in" might he have felt. The present data do not permit checking this interpretation.

Consistency of Evaluation and Yielding

After each series the subject answered a questionnaire concerning his reactions. The three items of present interest are:

1. Based on the previous nine judgments evaluate your ability to make these visual discriminations. Circle the number opposite the most appropriate evaluation.
2. Based on the previous nine judgments evaluate the ability of the others (on the average) to make these discriminations. Circle one.
3. Try to predict the average evaluation the others will make of your performance on the previous nine judgments. Circle one.

Each item was followed by the same six-point rating scale:

1. Very satisfactory (+3)
2. Quite satisfactory (+2)
3. Somewhat satisfactory (+1)
4. Somewhat unsatisfactory (−1)
5. Quite unsatisfactory (−2)
6. Very unsatisfactory (−3)

The subject's responses were weighted from +3 to −3 to reflect the direction and strength of the evaluation. For each subject on each questionnaire, we computed the degree of balance of his evaluation by multiplying the signed values for each item. Thus, the degree of balance in a triad of evaluations with values +2, −1, and −2 is +4. This procedure yielded 10 possible numerical values with either a positive or negative sign: 1, 2, 3, 4, 6, 8, 9, 12, 18, 27. The interval between adjacent values becomes larger as the values increase, although there is little basis for believing that the intervals between large values are psychologically larger. Increasing the value of one of the bonds by some amount should have roughly the same effect regardless of the initial level of balance. In order to approximate this state of affairs we performed a rank transformation on the balance values which we then treated as a metric scale assuming equal intervals. The 10 values, both positive and negative, were as-

TABLE 2
MEAN DEGREE OF CONSISTENCY FOR SERIES BY ERRORS

Errors	N	Series				
		I	II	III	IV	Total
0–1	23	3.00	3.39	4.52	4.74	3.91
2–4	24	0.13	2.29	3.25	3.25	2.23
5–10	24	1.58	2.50	.67	2.54	1.82
11–24	25	1.88	2.88	2.52	3.60	2.72
Total	96	1.64	2.76	2.72	3.52	

signed ranks from 1 to 10 in order of increasing magnitude. For example, a value of +18 now becomes +9 and a value of −27 becomes −10.

In examining balance tendencies over time we used the expedient of classifying the subjects as to the total number of errors for all four series. The maximum number of errors is 24 since there are six critical trials for each of the four series. Initially we examined the balance data on a series-by-series error rate, but the evaluation of differences became unwieldy. We therefore relied upon total errors as the basis for classification in order to obviate the difficulty encountered with those subjects whose error rate varied a great deal from series to series. The trends resulting from classification by total errors are in all essential respects similar to those based on a series-by-series error classification.

Table 2 presents the average degree of balance, as determined from the questionnaire items, by total errors and series. The error categories were chosen in order to obtain approximately equal numbers of subjects in each. We predicted that the subject would tend to reduce inconsistency over time. This is clearly evident in the column totals for each series. The variance due to series is significant at the .10 level; the mean square due to the linear component of the three degrees of freedom (*df*) for series is significant at beyond the .02 level.

There are certain obvious shortcomings of the balance metric used in the foregoing analysis. To what extent does a linear ordering of the 10 values violate the psychology of the situation? For example, does a +2, +2, +2 configuration possess two units more of balance than a +1, +2, +3 configuration? Offhand, this does not seem reasonable. Our metric no doubt does considerable violence to the data. More probably the 10 states represent some partial order. Short of a detailed investigation of the psychological space represented by these 10 configurations we relied upon our rough and ready metric. We decided, however, to use another, less powerful evaluation which makes fewer assumptions. We determined the proportion of consistent evaluation triads (those having a plus [+] value) for Series I and Series IV. The proportion for Series I is .62 (60/96) whereas for Series IV it is .80 (77/96) ($p < .05$ by chi square). We can accept this difference as adding some conviction to the conclusion reached via the detailed information offered by our questionable metric.

TABLE 3

NUMBER OF SUBJECTS IN EACH OF THE POSSIBLE PATTERNS OF CONSISTENCY AND INCONSISTENCY IN SELF-OTHER EVALUATIONS

	Series				*N*
	I	II	III	IV	
1	−	−	−	−	13
2	−	−	−	+	11
3	−	−	+	−	4
4	−	−	+	+	6
5	−	+	−	−	2
6	−	+	−	+	0
7	−	+	+	−	0
8	−	+	+	+	12
9	+	−	−	−	0
10	+	−	−	+	1
11	+	−	+	−	2
12	+	−	+	+	3
13	+	+	−	−	1
14	+	+	−	+	3
15	+	+	+	−	1
16	+	+	+	+	37

Our other prediction is also confirmed. The tendency toward balance in self-other attitudes is associated with behavior evidencing minimal conflict. Let us consider a subject with 0–1 errors as a consistent nonyielder, a subject with 2–4 errors as somewhat nonyielding, a subject with 5–10 errors as somewhat yielding, and a subject with 11–24 errors as yielding. For the two extreme categories, where there is presumably a strong commitment to the behavior, consistency tends to be greater than in the two middle categories. The variance due to errors is significant at beyond the .05 level. The quadratic component of the 3 *df* involved is significant at beyond the .02 level indicating the trend we predicted.

Again we performed the simpler analysis based upon the proportion of balanced triads. The two extreme categories (0–1 and 11–24) were combined and compared with the two middle categories (2–4 and 5–10). This dual classification represents a difference in the degree of commitment to behavior reflecting, we have assumed, a difference in conflict. The proportion of consistent triads over all four series in the extreme categories is .78 (150/192) and in the middle categories is .66 (127/192) ($p < .05$ by chi square). This difference over the four series reflects a consistent series-by-series trend.

TABLE 4

CHANGES IN EVALUATION COMPONENTS FOR SUBJECTS WHO INCREASE CONSISTENCY FROM SERIES I TO IV BY ERROR CATEGORY

Errors	Self-Evaluation			Evaluation of Others			Others Evaluation of Self		
	Increase	Decrease	Same	Increase	Decrease	Same	Increase	Decrease	Same
0–1	1	2	7	1	9	0	2	7	1
2–4	8	3	2	4	6	3	8	2	3
5–10	6	1	6	3	2	8	6	2	5
11–24	6	4	3	3	3	7	7	3	3

The increase in self-evaluation consistency can be seen more vividly with another type of analysis. Table 3 presents all 16 possible balance patterns over the four series. The first 8 patterns, which show initial imbalance, tend to become balanced. This effect can be tested by comparing the cell frequencies in the column at the far right for Patterns 2, 4, and 8, which are those that show a change toward balance, with the frequencies for Patterns 1, 3, 5, 6, and 7 ($p < .001$ by a binomial summation test). We obtain this large difference in spite of having included those subjects in Pattern 1 who may have changed to a balanced configuration had the experiment continued. The table also shows that those subjects who start out with a consistent self-evaluation (Patterns 9 through 16) tend to remain consistent ($p < .001$). Those few subjects who do change to an inconsistent configuration tend to change back again. Out of a total of 48 subjects who start out as consistent only 4 are inconsistent on Series IV.

One may inquire as to changes occurring over time in each of the three evaluation components for those subjects who show an increase in consistency. These data are shown in Table 4. The frequencies indicated are for an increase, a decrease, or no change in each of the evaluations in comparing Series I to Series IV for subjects who fall into the four error categories. The data for the subjects who did not change or who became less consistent are not shown. When the subject remained steadfast (0–1 category) he was clearly different from the others in his mode of increasing consistency. He tended to decrease both his evaluation of the others and the evaluation he believed the others had of him. This incidentally is coupled with an initial high self-evaluation. The other three categories, which are similar to each other, present a different picture. Here we see that the subject's mode of increasing balance is to increase both self-evaluation and the evaluation he assumes the other has of him.

We were also interested in determining the extent to which acceptance or rejection, i.e., degree of yielding, evidenced a diminution of conflict over time. There appears to be a progressive increase from series-to-series in the number of subjects who made either six errors or no errors, a trend that suggests that conflict did tend to decrease. To evaluate this change we compared the proportion of subjects who made six errors or none on Series I and on Series IV. The proportion on Series I is .29 (28/96) whereas on Series IV it is .47 (45/96) ($p < .02$ by chi square). We also compared for the first and last series, the proportion of subjects who made 0, 1, 5, or 6 errors. These proportions are .51 (49/96) and .75 (72/96), respectively, ($p < .01$ by chi square). This would seem ample evidence of conflict resolution over time. Table 5 shows the complete picture of series-by-series errors.

Questionnaire Evidence of Conflict

We made an effort to obtain the subjects' report of the degree of conflict he experienced during each successive series. We asked the subject to indicate on each questionnaire the number of judgments in the previous nine on which he believed that he experienced more difficulty than the others and the number of these judgments on which the average person would be perturbed. Neither question is direct but we believed that the subject would be less defensive in answering these indirect questions. The difficulty item shows a curvilinear trend

TABLE 5

NUMBER OF SUBJECTS IN EACH ERROR CATEGORY FOR EACH SERIES OF JUDGMENTS

Series	0	1	2	3	4	5	6
I	26	17	22	15	10	4	2
II	33	16	18	11	11	2	5
III	33	21	14	14	4	4	6
IV	38	23	12	10	2	4	7

which tends to mirror the consistency data. The responses were summed over the four series as were the errors. The trend is roughly summarized if we categorize total errors into our 0–1, 2–4, 5–10, 11–24 categories and examine the proportion of subjects in each category who claim to have had more difficulty on 10 or more judgments (of 36). The proportions are .17 for 0–1 errors, .36 for 2–4 errors, .63 for 5–10 errors, and .48 for 11–24 errors. We note a sharp increase through the 5–10 category followed by an abrupt decrease. The two successive increases are quite strong ($p <$ .01 for each by chi square). The decrease does not quite approach significance. However, if the decrease were evaluated against an extrapolated trend, the difference would be quite marked.

On the intended "projective" question, "on how many of the previous nine slides would the average person be perturbed," no such trend appeared. The data are arrayed in a triangular fashion. As the number of errors increase, the range of the number of projected perturbed responses tends to narrow. Trichotomizing the total number of perturbed responses into 0–7, 8–16, and 17–36, produces approximately equal n's in each category. If we examine the proportion of subjects falling into the middle category, i.e., having 8–16 perturbed responses, for the 0–1, 2–4, 5–10, 11–24 error categories these proportions are, respectively, .30, .46, .54, and .60. We note that the proportion doubles between the 0–1 and 11–24 error categories ($p < .05$ by chi square). It appears that among those subjects who did not yield were the most calm and most perturbed individuals.

These two questionnaire items do not appear to reflect the same thing. Projected conflict and estimated difficulty may not necessarily be highly correlated. Furthermore it is impossible at this point to know if the subject's response referred to the initial difficulty of the judgment, difficulty in retrospect after having made his choice, or a denial in an attempt to achieve congruence with his choice.

The perturbance question introduces the further contamination of any projective item: to what extent is the subject projecting his own psychological state and to what extent is he imagining how someone else would indeed react? To what extent does he consider himself average and in what sense does he interpret "average"?

A further distressing problem is created by the skewed distribution of judgment errors. In balancing the data for category comparison we required wider categories with increasing errors. The minimum error category is 1 error wide whereas the maximum category is 14 errors wide! Offhand, we would expect greater heterogeneity the wider the category. The perturbance data suggest the opposite. There is also a slight statistically insignificant trend for the 0–1 category to be most heterogeneous with respect to degree of consistency. Any reconciliation of this inverse relationship is beyond the present data.

The consistency and error data present a fairly clear-cut case for our hypotheses. The questionnaire data, if we reluctantly accept them at face value, suggest complexities that go much beyond the present data. We are currently pursuing a number of these problems.

SUMMARY

We explored some effects of future perspective with respect to group memberships upon the tendency of an individual to conform to the judgments of other group members when confronted with disagreement with them. Expectations were induced of future tasks that would be similar or dissimilar to the present one. Expectations of either continuing or noncontinuing group membership were also manipulated. There was a tendency to reject the others' judgments under both the anticipation of a continuous task and of noncontinuing membership.

We also found that the strength of the individual's tendency to yield to the others or

to remain adamant was correlated with his self-other evaluation consistency.

REFERENCES

ASCH, S. E. Effects of group pressures upon the modification and distortion of judgments. In M. H. Geutzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie Press, 1951. Pp. 117–190.

FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. H. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327–346.

GERARD, H. B. Some determinants of self-evaluation. *J. abnorm. soc. Psychol.*, 1961, **62**, 288–293.

HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1955, **50**, 363–371.

(Received May 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 573-577

FAMILIARITY AND THE ORDER OF PRESENTATION OF PERSUASIVE COMMUNICATIONS[1]

ROBERT E. LANA
American University

CARL Hovland and his associates (Hovland, 1957; Hovland, Janis, & Kelley, 1953) have developed a program of research in the field of communication and persuasion during the last few years that represents the most comprehensive experimental and theoretical work in this field to date. One of the most recent books by Hovland, *The Order of Presentation in Persuasion*, includes a compendium of research examining various methodological and conceptual questions concerning the position of various types of affirmative and negative arguments and their individual and combined effects on attitude and opinion change. One of the persisting problems which Hovland recognized, but did not directly examine, is that concerning previous familiarity of the subject with the topic of communication and the corresponding effect of the order of presentation of conflicting arguments on attitude or opinion change. Two early key studies by Lund (1925) and Cromwell (1950) indicated that neither a primacy nor a recency effect is immediately evident in the situation where two communications on a single topic, one affirmative and the other negative, are presented successively. Primacy refers to the situation where the argument given first, whether pro or con, is more effective in changing opinion in the direction of the argument. Recency refers to a similar effectiveness of the argument presented second. In Lund's study arguments were presented successively for and against an issue, affirmative arguments first for half the group, and negative arguments first for the other half. Pretest questionnaires were filled out before the first communication and again after the first and second communications. The change resulting from the first communication was greater than that resulting from the second. This was described by Lund as the law of primacy in communication. Cromwell later reported directly opposite results where speeches presented in the second position had a greater influence on the attitude of the listeners than those presented first. In this latter study the effects were evaluated after the two sides had been presented. In the Lund study effects were evaluated after each side had been presented. In replicating both of these studies Hovland and Mandell found no consistent primacy or recency effects. Among the various experiments discussed by Hovland (1957), attempting to determine the various factors influencing primacy and recency effects, none deal directly with the problem of previous familiarity of the subjects with the topic of the communication. However, Hovland and his coauthors recognized the need for research in this direction. In the summary of their book, they state, "The results of the present series of studies seem to suggest that the nearer one comes to achieving primacy in the sense of the first presentation of unfamiliar material, the more apt one is to obtain primacy effects" (p. 139). Also, "Undoubtedly further research bearing directly on the relationship between prior familiarity with the topic under discussion and the relative effectiveness of the side of the issue presented first would greatly help to clarify the extent of generality of a law of primacy" (p. 140).

This paper is a report of research designed to investigate the effects of the order of presentation of a pro and con communication on opinion change of adult subjects who have been differentially familiarized with the topic of the communications. The greatest difficulty in the construction of the experimental design was

[1] This research was supported by the National Institute of Mental Health, United States Public Health Service Grant M3774 (A) to the American University on which the author served as principal investigator.

The Vivisection Questionnaire, the pro and con communications, and the long and short familiarization talks utilized in this study have been deposited with the American Documentation Institute. Order Document No. 6765 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

the necessity of insuring minimal initial familiarity of the subjects with the topic utilized. Since college students were to act as subjects, the opinion areas such as ethnic prejudice, political preference, etc. usually tapped in studies such as this one could not be used since all subjects would most likely be familiar with these topics a certain large but indeterminable degree. What was needed was a topic about which the subjects had little information and had formed little previous opinion. This topical area would also have to be uncomplicated enough so that the subjects could easily form an opinion when confronted with conflicting arguments relevant to it. In other words, topics such as those concerned with technical problems in constructing an H-bomb, etc. could not be used. With this consideration, the topic of animal vivisection was chosen as the opinion area to be used in this study because it is a topic that is only rarely considered in public discussions through the mass media (at least at present). It is a simpler matter to manipulate subjects' familiarity toward vivisection than it would be if a more popular or controversial topic were used. That is, initial familiarity toward vivisection should be close to zero for most people. An attempt to examine this assumption was made in the study.

METHOD

Materials

Two questions, "Do you know what the word vivisection means?" and "If you do describe its meaning in one or two sentences," were duplicated on sheets of paper and all subjects were asked to respond to them. A reliable 10-item Likert-type vivisection questionnaire, developed by Molnar (1955), was employed as the pretest and posttest in order to evaluate opinion toward vivisection along a continuum of favorability. Two informational standardized talks on vivisection, with no direct positive or negative statements about the practice, were administered to two of the groups. The first talk was 213 words in length and the second 92 words in length. The amount of information presented was roughly proportional to the number of words in each talk. A pro and a con argument, both of approximately 500 words length, concerning the topic of vivisection were used as the persuasive materials in the study. These arguments were developed by Molnar (1955) for use in another context and take 4 minutes to present. These same materials were used by the investigator to examine the related pretest-treatment interaction effect (Lana, 1959) and have proved to be excellent devices for that purpose.

TABLE 1
EXPERIMENTAL DESIGN

A_1		A_2		A_3	
Two questions Pretest		Two questions Pretest		Two questions Pretest	
Long familiarization talk and discussion		Short familiarization talk			
12 days		12 days		12 days	
B_1	B_2	B_1	B_2	B_1	B_2
Pro	Con	Pro	Con	Pro	Con
Con	Pro	Con	Pro	Con	Pro
Posttest		Posttest		Posttest	

TABLE 2
SUMMARY OF PRETEST AND POSTTEST MEAN SCORES AND DIFFERENCE SCORES

Score	A_1 (Very familiar)		A_2 (Familiar)		A_3 (Unfamiliar)	
	B_1 (Pro 1st)	B_2 (Con 1st)	B_1 (Pro 1st)	B_2 (Con 1st)	B_1 (Pro 1st)	B_2 (Con 1st)
Pre	35.08	35.90	35.88	36.00	36.73	36.38
Post	38.42	37.50	38.31	37.56	37.53	38.69
Gains	3.34	1.60	2.43	1.56	.80	2.31
(N)	(12)	(10)	(16)	(16)	(15)	(16)
Difference of differences	+1.74 (primacy)		+.87 (primacy)		−1.51 (recency)	

Experimental Design

A total of 85 undergraduates in three Psychology I classes at the American University served as subjects in the experiment. The subjects were randomly assigned to one of six treatment groups. All subjects were asked to answer the two questions concerning the nature of vivisection described above on a sheet of paper. After these papers were collected all subjects responded to the 10-item Vivisection Questionnaire which was also collected. Group A_1 was then read the long familiarization talk mentioned above and discussed the nature of vivisection with the experimenter until questions and comments from the class stopped. The discussion lasted approximately 10 minutes. The experimenter carefully avoided making any value judgments about the topic during the discussion. The group was then dismissed. Twelve days later this group was subdivided into two groups, one of which (B_1) listened to the pro communication and then, immediately following, the con communication. The other subgroup (B_2) listened to the two statements in reverse order, first con and then pro. Both were then posttested with the identical questionnaire that was used as the pretest. Group A_2 having answered the initial questions and responded to the questionnaire for the first time, listened to the shorter familiarization talk immediately afterward but was allowed no discussion. Twelve days later the group was

TABLE 3
RESULTS OF t TESTS

	Groups		
	A_1	A_2	A_3
Difference of differences	+1.74	+.87	−1.51
t[a]	2.03	1.44	2.11
p	<.05	>.05	<.05

[a] Combined error term with 79 df utilized in t tests.

subdivided into two groups (B_1 and B_2) and followed the same procedure as Group A_1. Group A_2 answered the questions and filled out the questionnaire only. Twelve days later the same procedure utilized with Groups A_1 and A_2 was carried out. The design is summarized in Table 1.

RESULTS

The responses to the two questions on the meaning of the term "vivisection" indicated that 95% of all subjects have little or no idea what the term means. Since the term vivisection has sometimes been used to imply a negative value judgment in occasional newspaper campaigns against the practice of surgical operation on live animals, it is probably not an unfamiliar term to individuals who are aware of the phenomena about which it refers. It is thus concluded that the large majority of subjects used in this experiment knew very little about the topic of vivisection and hence probably had formed, at most, only mild opinions about the matter. The experimental results were analyzed using a subtractive procedure described by Hovland and Mandell (Hovland, 1957). Pretest means for all groups are subtracted from their respective posttest mean scores. Each B_1 difference score is then subtracted from its appropriate B_2 score yielding three "difference between differences" scores, one for each of the three A groups. These final difference scores were then tested against the null hypothesis by utilizing a t test. If the mean differences are positive then a primacy effect has been demonstrated (if statistically significant). If the mean differences are negative, a recency effect has been shown (if statistically significant). A summary of the mean pretest and posttest scores for all groups and a summary of the analytic procedure are included in Table 2.

A Bartlett's test for homogeneity of variance was run on all scores and χ^2 was not significant at the .05 level. It was concluded that the scores were homogeneous with respect to variance. A combined error term for the posttest scores was then calculated and employed in the t tests involved in testing the three pairs of difference scores. A significant primacy effect was found in the "Very familiar" group and a significant recency effect in the "Unfamiliar" group. The "Familiar" group indicated a primacy effect which was not statistically significant. The .05 level of significance was set as the alpha level for the study. These results are summarized in Table 3.

DISCUSSION

The results demonstrate that increased familiarity produces increased primacy effects, at least under the conditions of this study. A recency effect is apparent in the group that was unfamiliar with the topic before the presentation of the conflicting communications. It is likely that the informational talks on vivisection influenced opinion as well as the persuasive arguments, perhaps in terms of increasing stability rather than in influencing the means of the various groups. However, care was taken to make the familiarization talks as neutral as possible. In any case, the major thesis of this paper is not disturbed by this possibility. Thus it has been shown that a communication presented first, whether pro or con, is more effective the more familiar the subject is with the topic. It should be noted, however, that the range of familiarity employed in this study probably extends from somewhere near zero to "somewhat familiar" at most. If a topic with which the subjects are very familiar, such as racial prejudice, were used, the degree of familiarity with the topic of the various experimental groups would vary from a base very much greater than zero to a very great familiarity compared with the "unfamiliarized" group in this study, which is close to zero familiarity. Perhaps different arrangements of primacy and recency effects for the various groups might result under these conditions. If, however, familiarity is greater than zero then primacy is not actually achieved in the sense of the first presentation of material. It may be that measurable primacy and recency effects only appear when one is dealing with a topic with which the subject is only slightly familiar or not familiar with at all.

Hovland and Mandell's (Hovland, 1957) study, which showed no clear cut primacy or recency effects using a topic with which the subjects were already familiar, and the results of the current study seem to support this contention.

A study by Luchins cited by Hovland (1957) is also relevant in this context. Luchins presented two different descriptions of the same individual to high school and college students. These descriptions were characterized by being polar in nature on an introversion-extroversion continuum, hence they are essentially opposed communications. The subjects read the descriptions from a sheet of paper. Complete lack of familiarity with the subject matter was assured since the character who was being described ("Jim") was fictitious. The subjects were then asked to describe Jim in their own words and answer questions about his personality characteristics. The presentation of the opposing descriptions were juxtaposed for the various experimental and control groups. Luchins found consistent primacy effects. He favors a "set" explanation of his results in that the initial description directed the latter opinions of the subjects in much the same way that a particular initial solution to a problem effects later attempts to solve similar problems.

Luchins' experiment is the only one in the literature bearing directly on this problem. All of the groups in his study began with essentially zero familiarity toward the topic. In the present study there are three levels of familiarity operating, and yet the presence of primacy or recency is opposite to the effects obtained by Luchins. That is, with increasing familiarity primacy increases, and recency occurs with low familiarity.

Luchins suggests certain possible reasons that may account for the discrepancies. The communications used in both studies were markedly different. In one case the subjects read the communication and in the other it was read to them. Luchins' communication describes a person, the vivisection communication a process. Any of these factors may be central in influencing the predominance of primacy or recency effects in a given communication situation. Perhaps major differences between Luchins' tasks and the present one could have accounted for the opposite results. In Luchins' study the subjects, being asked to describe an individual in terms of conflicting characteristics, were probably not set to weigh both possibilities of interpretation as they might be expected to weigh the two sides of an argument with respect to a social question such as vivisection. When the final questionnaire was administered in this study, it is probably safe to assume that the subjects realized they were expected to make a choice between the two major possibilities suggested by the pro and con vivisection arguments. In Luchins' study it is not probable that the subjects perceived themselves as being in a similar kind of choice situation. Since there have been relatively few experiments on the problem of familiarity and its effect on order of presentation of a persuasive communication, and since the differences among these studies may be attributable to technical and methodological difficulties rather than theoretical ones, any attempt at a theoretical explanation of the results of this study would be gratuitous.

The problems just raised indicate the need for further research before any meaningful theoretical interpretation of primacy, recency, and familiarity can be attempted. The indication of a recency effect when the topic of a communication is unfamiliar to the recipient and an increasing tendency for primacy with increasing familiarity shown in this study must be examined in context with several other variables yet to be studied before any theoretical conclusions can be made. Recency effects have been demonstrated by other workers in various situations which must be considered in any attempt to attain understanding of the total problem of order of presentation. Luchins (1958) found a recency effect when a questionnaire was distributed after each of the two opposing communications had been delivered when the topic included different descriptions of a single individual. Miller and Campbell (1959) showed that recency effects were strongest when there was a long delay between first and second communications with immediate measurement after the second communication. When the two arguments are presented successively and the testing delayed for a considerable time, recency is at a mini-

mum. Film, tape recorded, written, and spoken presentation of persuasive arguments must be examined to determine any possible differential effects of these variables on opinion change. Communications about controversial topics such as racial prejudice or political problems may have different effects on order of presentation and familiarity factors. The question as to whether the subject is aware of the intent of the communicator must also be answered. Some interview or questionnaire technique would have to be developed for this purpose. Before studies attempting to solve some of these problems are completed, anything approaching a comprehensive theoretical formulation of communication effects on opinion will be extremely difficult.

SUMMARY AND CONCLUSIONS

This study was designed to examine the influence of the order of presentation of a pro and a con communication about vivisection on adult subjects who have been differentially familiarized with the topic of communication.

Eighty-five subjects were divided into three experimental groups which received a long familiarization talk and discussion on vivisection, a short familiarization talk with no discussion, and neither a talk nor discussion, respectively, after all groups had answered two questions as to the meaning of the term "vivisection" and had filled out a vivisection opinion questionnaire. After 12 days each group was divided into two subgroups, one of which received a pro and then a con argument about vivisection, and the other of which listened to the arguments in reverse order. All groups then responded to the vivisection opinion questionnaire for the second time. An analysis of difference scores between pre and posttest mean scores for all groups indicated that the argument in first position (primacy effect) was most successful in changing opinion for the group with the greatest familiarization. The argument in second position (recency effect) was most effective in changing the opinion of the group with no familiarization.

It is concluded that:

1. Under the conditions of the experiment, increasing familiarity with the topic of a communication on the part of the subjects is directly related to an increasing primacy effect when opposing arguments are presented successively.

2. A recency effect is evident when the subjects' familiarity with the topic of communication is near zero and when opposing arguments are presented successively.

REFERENCES

CROMWELL, H. The relative effect on audience attitude of the first versus the second argumentative speech of a series. *Speech Monogr.*, 1950, **17**, 105–122.

HOVLAND, C. I. (Ed.) *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion*. New Haven: Yale Univer. Press, 1953.

LANA, R. E. Pretest-treatment interaction effects in attitudinal studies. *Psychol. Bull.*, 1959, **56**, 293–300.

LUCHINS, A. S. Definitiveness of impression and primacy-recency in communications. *J. soc. Psychol.*, 1958, **48**, 275–290.

LUND, F. H. The psychology of belief: IV. The law of primacy in persuasion. *J. abnorm. soc. Psychol.*, 1925, **20**, 183–191.

MILLER, N., & CAMPBELL, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59**, 1–9.

MOLNAR, A. The effects of styles, speakers, and arguments on the attitudes and perceptions of a listening audience. Unpublished master's thesis, University of Maryland, 1955.

(Received May 13, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 578-585

OPINION EVALUATION AND AFFILIATION[1]

ROLAND RADLOFF[2]
University of Minnesota

THE present study was designed to test the hypothesis that a person's needs for self-evaluation regarding the correctness of his opinions arouse affiliative tendencies in him. It stems directly from results and discussion contained in Schachter's (1959) recent book *The Psychology of Affiliation* and is related to Festinger's (1954) theory of social comparison processes.

Schachter (1959) found a positive relation between emotional arousal and affiliative tendencies; specifically, under conditions of experimentally manipulated anxiety, highly anxious subjects wanted to be with others more than did nonanxious subjects. From a series of experiments designed and executed to explore the determinants of this relationship, Schachter concluded that the affiliative tendencies he found were a manifestation of "needs for anxiety reduction and needs for self-evaluation; that is, ambiguous situations or feelings lead to a desire to be with others as a means of socially evaluating and determining the 'appropriate' and proper reaction" (p. 132).

Festinger (1954) has postulated that: "There exists, in the human organism, a drive to evaluate his opinions and abilities" (p. 117). Further, Festinger proposes that in the absence of objective or physical evaluative criteria, opinions are evaluated by comparing them with those of other people. The opinion is then taken as correct or incorrect in terms of agreement or disagreement with the opinions that the others hold. The drive for self-evaluation has been successfully employed to account for a variety of behaviors in a number of laboratory studies. Specifically, group members have been found to change their opinions, (Back, 1951), exert influence to change the opinions of others (Festinger, Gerard, Hymovitch, Kelley, & Raven, 1952), and reject unchanging deviates in order to create uniformity of opinion within the group (Schachter, 1951), thereby providing validity for opinions by consensus. In all of the studies just cited, opinion comparisons were made in existing groups. No experimental evidence is yet available that bears directly on the question of opinion evaluation and affiliation. The present study was therefore planned to demonstrate that needs for evaluation of opinions are important determinants of group formation in addition to their function as sources of pressure to uniformity within groups. The major hypothesis of the study is: A person who is uncertain about the correctness of one of his opinions, for which he finds no objective criteria available by which to evaluate its correctness, should seek affiliation with other people in order to evaluate his opinion via social comparison.

In an experimental test of this hypothesis, specific predictions regarding affiliative tendencies would follow from a manipulation of evaluative needs. With initial need arousal held constant, if for some subjects information that satisfies evaluative needs is provided while for others it is not, affiliative tendencies should be stronger in the latter group than in the former. Since people are assumed to satisfy evaluative needs by comparing their opinions with those of others, one method of manipulating evaluative information would be to provide subjects in different conditions with information about the opinions of various groups of other people. Such information can be ordered on a continuum of adequacy for purposes of comparison and evaluation. Points ranging from inadequate to adequate could be represented by giving the subjects information about the opinions of: (*a*) no others, (*b*) irrelevant others or inferiors, (*c*) peers, (*d*) ex-

[1] This study is part of a dissertation (Radloff, 1959) submitted to the faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This study is also part of a program of research on affiliation behavior that is being conducted at the Laboratory for Research in Social Relations at the University of Minnesota. The program is under the direction of Stanley Schachter and is supported by a research grant from the National Institute of Mental Health, United States Public Health Service.

[2] The author is now with the United States Public Health Service.

The author wishes to thank Stanley Schachter for support and advice on all phases of the study.

perts. (Inferiors are defined as people who have less education, information, and experience than the subjects; and whose inferiorities are relevant to the validity of their opinions on the issue in question. Experts are defined as people who have greater education, information, and experience than the subjects, and whose superiorities are relevant to the issue in question.)

The magnitude of evaluative needs (and therefore of affiliative tendencies) are predicted to vary in the following manner according to the conditions just specified $a > b > c > d$. These predictions depend in part on the degrees of inferiority and expertness that the subjects attribute to groups b and d. According to the strength of the manipulation and the subject's perception, b could be closer to a or to c, and d could be close to or distant from c in terms of adequacy as comparison groups. In the absence of prior knowledge regarding the strength of the manipulations, predictions can be made only for the rank order of the groups and not for the magnitude of the differences.

The manipulations just considered assume that the subjects' commitment to an opinion is held constant. If commitment is allowed to vary a person committed to an opinion may be expected to have greater evaluative needs than someone who is not so committed.

Differences can be predicted within the groups according to the subjects' agreement or disagreement with the opinions of the others provided him. According to Festinger (1954), opinion evaluation is impaired when the only comparison available is a very divergent one. Therefore, subjects in the peer comparison condition who disagree with the opinions of a large majority of the comparison group should have evaluative needs that are relatively unsatisfied as compared with those who agree with the opinions of the large majority of the comparison group. Evaluative needs should be less effected by agreement or disagreement with a large majority of the members of inferior or expert comparison groups than they are by agreement or disagreement with peers. The magnitude of such differential effects is again dependent upon the extent to which these groups are perceived to be different from the peer group. Information about the opinions of inferiors should be irrelevant for both those who agree and those who disagree if the inferiors are sufficiently low in status. Receipt of information regarding the opinions of experts provides adequate evaluative information for both those who agree and those who disagree if the experts are sufficiently high in status.

Further within-group differences may be predicted on the basis of individual differences. Schachter (1959) found an interaction between birth order, anxiety, and affiliation. Specifically, he found that affiliation tendencies of early born subjects were more strongly effected by the manipulation of anxiety than were those of later born subjects. Schachter attributed these differences to the greater dependency needs of first born and only children. Relevant to the present study the question may be asked, is this effect specific to anxiety or does it apply to a variety of situations in which the need for social evaluation is a motive? A partial answer to this question can emerge from a test of the following hypothesis in the present study: First born and only subjects, who are presumably more dependent on other people for the evaluation of their opinions than are later born subjects, should manifest greater differences in affiliative tendencies as a result of varying needs for social evaluation.

METHOD

The experiment designed to test these hypotheses was conducted as a combined survey of student opinion and a recruiting session for participants in discussion groups. Subjects were first asked to give their opinions on an issue in writing. Opportunities for evaluation were manipulated by telling them the opinions of groups of other people. Affiliation tendencies were measured by asking the subjects to volunteer for discussion groups that were to talk about the issue on which they had just given their opinions.

Procedure

Subjects were 480 females from introductory psychology classes at the University of Minnesota, who were randomly assigned to experimental groups within the time limits their schedules would permit.

Subjects were assembled in groups of 25–30. After they had filled out a short questionnaire giving data on birth order, the experimenter read to all groups a statement designed to arouse interest in the topic, impress upon subjects the importance of the opinion they were asked to give, and provide a rationale for the procedure to follow. The gist of the remarks was that the number of potential students and costs of college educa-

tion would present serious problems by 1970. Data were given to emphasize the gravity of the situation. The specific aspect of the problem the subjects were asked to consider was, who shall pay these increased costs, students or governmental units? Their stake in the problem as future parents and taxpayers was emphasized, and they were told that their special information on and concern for the problem would add weight to their opinions.

After hearing the statement, each subject received an opinion survey form asking her to "Please indicate by a check mark on the following scale what per cent of the cost of their education you feel college students or their parents should pay." The scale ranged from 0% to 100% with intervals of 5%. Subjects were also asked to indicate how certain they were of their opinions on a seven-point scale with categories ranging from "absolutely certain" to "completely uncertain."

Manipulation of Evaluative Information

At this point subjects were provided with differential opportunities to evaluate the correctness of their opinions by giving them information about the opinions of others, in the form of a frequency distribution that was drawn on the blackboard and explained in detail to the subjects. The abscissa of the distribution was the same scale on which subjects had made a check of their opinions; the ordinate represented frequency of response in the particular comparison group. For all groups given information the curves were identical in shape, skewed with a highly peaked mode. As a control for possible interaction between the subject's position on the opinion scale and agreement or disagreement with others, half the groups in each condition were shown a distribution with the mode between 20% and 40%, while the other half were shown a distribution with the mode between 60% and 80%. Seventy-five percent of the responses of the comparison group were ostensibly in this 20% range, with 12% in the long tail and 13% in the short tail of the distribution. By means of this graph each subject could see that either an overwhelming proportion of the comparison group (75%) agreed with her, or that an even larger proportion (87% or 88%) disagreed with her, according to the position she had checked on the opinion survey form.

These manipulations produced the following conditions:

No Information. Subjects in this condition were not given any information about the opinions of other people. That is, no frequency distribution was shown.

Inferior. Subjects in this condition were shown a distribution purportedly summarizing the opinions of high school sophomores.

Peer. Subjects in this condition were shown a distribution purportedly summarizing the opinions of college sophomores at their university.

Expert. Subjects in this condition were shown a distribution purportedly summarizing the opinions of leading economists, university presidents, and state legislators concerned with financing higher education.

No Decision. In a fifth condition, the extent to which the subjects held an opinion on the issue was manipulated by omitting administration of the survey form. Subjects in this condition were not given any information about the opinions of other people. Their interest in joining a discussion group was solicited immediately after the interest arousing statement had been read.

Measuring Affiliative Tendencies

Affiliation tendencies were measured by asking subjects to volunteer to take part in discussion groups that were to talk about the problem of financing higher education. The experimenter stated that the discussion groups would be composed of six to eight "students like yourselves." The ostensible purpose of the discussion groups was to "generate some original ideas that can be put to use." An information form was distributed on which the subject was asked to give her name and address and to indicate the amount of interest she had in volunteering for a discussion group. The four choices available were:

1. I would like very much to take part in a discussion group.
2. I wouldn't mind taking part in a discussion group.
3. I could take part in a discussion group if necessary.
4. I have no interest in taking part in a discussion group.

This instrument provided the measure of the dependent variable.

Independent Measure of Need for Social Evaluation

Subjects were asked their reasons for volunteering to join discussion groups as a check on the need for social evaluation. An open-ended questionnaire was distributed with the explanation that the Student Opinion Research Center, sponsors of the study, conducted many such discussion groups, and was interested in the problem of why people do and do not take part in such groups in order that more effective recruitment techniques might be employed. Since this questionnaire was not introduced until the experiment had been in progress for some time, data are available for only a portion of the subjects.

At the conclusion of the experiment subjects were asked not to discuss the issue with other psychology students since the experimenter wanted to have a true expression of student opinion, uncontaminated by information about how others felt. A questionnaire labeled "Previous Interest" revealed that subjects in later groups had had no discussions with those in previous groups, and that they did not suspect the true nature of the experiment.

RESULTS

Differences in Affiliation Tendencies between Conditions

The major hypothesis of a positive relation between need for opinion evaluation and

TABLE 1
MEANS OF INTEREST IN JOINING A DISCUSSION GROUP AND PERCENTAGE IN EACH GROUP INDICATING MAXIMUM INTEREST

Condition	N	Mean Interest in Joining a Discussion Group[a]	Percentage Indicating Maximum Interest
No Information	114	2.02	40%
Inferior	93	2.30	30%
Peer	107	2.47	19%
Expert	87	2.63	14%
$F = 6.98$; $df = 396,3$; $p < .001$			
No Decision	79	2.53	12%

[a] Lower scores indicate greater interest in joining discussion groups.

affiliation tendencies is strongly supported by the data reported in Table 1. Means indicating the strength of affiliation tendencies are in the predicted order of magnitude for the four conditions and the variance of the condition means around the grand mean is highly significant. The exact probability of the four means being ranked in the predicted order is .042 since there are 24 possible ways in which they could have been ordered by chance. Thus it can be concluded that providing subjects with differential opportunities to evaluate their opinions strongly effected affiliative tendencies.

A comparison of the No Information and No Decision condition affiliation means reveals that holding an opinion also effects affiliative tendencies. Subjects who were asked to form and record their opinions exhibited significantly stronger affiliative tendencies than those who were not asked to do so.

Effects of Agreeing or Disagreeing with Others

According to the point they had checked in recording their opinions, subjects were divided into two groups for the purpose of an analysis of differences within conditions. If the check mark indicating the subject's opinion lay within the 20-point modal range of the comparison distribution that she had been shown, the subject was classified as a "conformer"; if it lay outside this range, she was classified a "deviate."

Results presented in Table 2 show no significant differences between the affiliation tendencies in any of the three conditions in which information was provided. However, a straight "conformer-deviate" division assumes that the subjects who were deviates at the time their opinions were measured were also deviates at the time affiliation tendencies were measured. Festinger's (1954) theory of social comparison processes holds, and numerous experimental studies (Back, 1951; Festinger et al., 1952; Festinger & Thibaut, 1951; Gerard, 1953; Hochbaum, 1954) have found, that one manifestation of pressures toward uniformity is a change of opinion toward the group norm by deviates. It may therefore be presumed that in the present study, many subjects who were deviates according to their opinion statements may no longer have been deviates at the time their affiliation tendencies were measured, having changed their opinions to conform to the norm of the comparison group. Since it was not feasible to include a direct measure of opinion change in the study, the measure of certainty of opinion was employed to separate those who presumedly changed their opinions from those who did not. In previous studies in which opinion change has been measured, Hochbaum (1954) and Ehrlich (1958) found that subjects who were uncertain of their opinions changed to conform to a normative opinion significantly more than did those who were certain. Therefore, in the present study, subjects were dichotomized near the median according to their responses to the scale measuring certainty of opinion. Subjects who indicated that they were "absolutely," "very," or "quite" certain were called *certain*; while subjects checking "fairly" and "not at all" certain and "very" and "completely" uncertain were designated *uncertain* for purposes of analysis.

Comparisons of the affiliation means reported in Table 3 give strong support to the hypothesis that deviates in the Peer condition who were certain of their opinions have unsatisfied evaluative needs. The affiliation tendencies of subjects in this group were nearly

TABLE 2
MEAN AFFILIATION SCORES FOR CONFORMERS AND DEVIATES

Condition	Conformers	N	Deviates	N
Inferior	2.19	27	2.35	66
Peer	2.53	34	2.44	73
Expert	2.70	23	2.61	64

TABLE 3
MEAN AFFILIATION SCORES FOR CERTAIN AND UNCERTAIN DEVIATES

Condition	Deviates Who Were Certain of Their Opinions	*N*	Deviates Who Were Uncertain of Their Opinions	*N*
Inferior	2.24	34	2.47	32
Peer	2.09	32	2.71	41
Expert	2.54	35	2.69	29

equal to the affiliation tendencies of subjects in the No Information group. Peer deviates who were certain of their opinions manifested stronger affiliative tendencies than did Peer conformers ($p < .10$) or Peer deviates who were uncertain of their opinions ($p < .01$). Also, as predicted, there were no differences within either the Inferior or Expert conditions between deviates who were certain of their opinions and deviates who were uncertain of their opinions and conformers.

Analysis of Subjects' Reasons for Joining Groups

Data presented in Table 4 provide a direct measure of the inferred construct "need for social evaluation." To simplify the analysis for "reasons" and birth order, subjects were divided into two groups—one group predicted to have *high needs* for opinion evaluation and the other to have *low needs*, according to the theory of social comparison processes. Needs for evaluation were expected to remain high for subjects in the No Information and Inferior conditions and for deviates in the Peer condition who were certain of their opinions; they should be low for subjects in the Expert condition, for deviates who were uncertain of their opinions and conformers in the Peer condition. Affiliation means for the "High Need" and "Low Need" groups, respectively, were 2.14 and 2.63 ($p < .001$). The *N*s in Table 4 are quite small compared to these in previous tables, since these data were gathered on only a portion of the subjects in each condition and very few subjects expressed interest in joining discussion groups in the "Low Need" group. Reasons were given in response to the open-ended request: "Please indicate why you are or are not interested in joining a discussion group to talk about this issue." Each subject was free to give as many reasons as she desired, but the mean number of reasons given by the High Need and Low Need groups and the proportional frequencies of subjects stating a specific number of reasons were identical. A reason was coded as expressing a need for social evaluation only if the subject mentioned wanting to hear the views of other people or wanting to compare her views with those of others. Other categories of reasons for wanting to join discussion groups were: social or "fun" aspects of groups, interesting nature of discussion groups, information seeking (no mention of getting ideas from other participants), egocentric—wanting to make certain their opinion was heard, wanting to help own financial position, wanting to help younger sibs or their own children, performing a duty for society.

The data support the hypothesis that needs for social evaluation account for differences in affiliation tendencies between the two groups. Among subjects who desired to join discussion groups, those in the High Need group gave reasons indicating needs for social evaluation significantly more often than did those in the Low Need group ($p < .05$). In other words, the greater affiliative tendencies of subjects in the High Need group may be directly attributed to greater needs for social evaluation.

The classification in terms of High Need and Low Need divides subjects from the Peer condition on theoretical grounds. While this internal analysis was necessary in order to provide an *N* of sufficient magnitude to permit a reliable measure of the "reasons" index, similar differences with respect to "reasons" apply as well for the treatment groups that could be classified in their entirety as "High" "Low" in need. Thus subjects from the No Information and Inferior conditions who de-

TABLE 4
SOCIAL EVALUATION REASONS VERSUS OTHER REASONS FOR WANTING TO JOIN DISCUSSION GROUPS[a]

Group	Social Evaulation Seeking Reasons	Other Reasons	Total
High Need	20	39	59
Low Need	3	21	24
Total	23	60	83

$\chi^2 = 3.89$, $df = 1$, $p < .05$

[a] Two coders agreed on the assignment of reasons to categories in 100% of the cases.

sired to join groups gave reasons indicating needs for social evaluation more frequently than did their affiliatively inclined counterparts from the Expert condition ($.10 > p > .05$).

Individual Differences: The Effects of Birth Order

Data presented in Table 5 indicate that, in support of the hypothesis concerned with birth order, manipulation of evaluative needs produced a large and significant difference in affiliation tendencies for early borns ($p < .001$) and a smaller, nonsignificant difference for later borns. The interaction between birth order and needs is significant ($p < .01$). The results of the present experiment complement Schachter's (1959) findings that early and later borns do not differ in overall affiliative tendencies; they only differ when dependency-related needs are aroused.

Again, it should be noted that while the analysis according to birth order includes subjects from the Peer condition, differences of similar magnitude are obtained when only the conditions that could be assigned entirely to "High" or "Low" need are considered (Table 6). Although the interaction of birth order with need is less reliable ($.05 > p > .01$) in Table 6 than it is in Table 5, the mean affiliation scores are virtually identical in the two tables.

An examination of the reasons for joining discussion groups provides evidence that the greater affiliative tendencies of early borns in the High Need conditions were motivated by needs for social evaluation. Among subjects from whom reasons were solicited, there were 17 early borns and 14 later borns who showed maximum interest in joining groups in the High Need conditions. Of these subjects, 9 early borns and 3 later borns gave reasons indicating need for social evaluation. The exact probability of a distribution this extreme or more so occurring by chance, given fixed marginals, is .075.

TABLE 5
AFFILIATION MEANS BY BIRTH ORDER AND CONDITION

Birth Order Condition	Only and First	*N*	Later	*N*	Row Means
High Need	2.03	118	2.26	116	2.14
Low Need	2.81	74	2.50	82	2.65
Column Means	2.33		2.36		

TABLE 6
AFFILIATION MEANS BY BIRTH ORDER AND CONDITION

Birth Order Condition	Only and First	*N*	Later	*N*	Row Means
No Information and Inferiors	2.00	99	2.28	108	2.14
Experts	2.83	35	2.50	52	2.63
Column Means	2.22		2.35		

DISCUSSION

The results of the present experiment support the hypothesis that the drive for self-evaluation with respect to opinions can be an important determinant of affiliation.

As is generally the case in experimental social psychology, the manipulations in the present study were relatively weak. One could therefore not expect to arouse affiliative tendencies in all subjects in the High Need group. The contention is plausible, however, that most people at some time experience evaluative needs which can be satisfied only by social comparison, of sufficient strength to motivate them to seek out others. The relationship established would thus seem to have wide generality.

A theoretically significant contribution of the present study is the demonstration that the affiliative consequences of evaluative needs, first employed by Schachter (1959) to account for affiliative responses under conditions of emotional arousal, are apparently applicable in a variety of seemingly disparate contexts. That is, whenever evaluative needs capable of being satisfied by social comparison are aroused, the arousal of affiliative tendencies is a potential consequence. The arousal of evaluative needs concerning abilities, for example, should also produce affiliative tendencies in the absence of adequate nonsocial means of evaluation.

No Decision versus No Information

The large difference in affiliation tendencies between subjects in the No Information and No Decision conditions raises the question, what are the consequences of stating an

opinion? One probable consequence is an increase in the salience of the issue; thus subjects in the No Information condition may have a greater desire to discuss an issue which is more important or interesting to them. Another possible consequence is the generation of cognitive dissonance (Festinger, 1957) which may arise when an individual commits himself to a position on a complicated, multisided issue. In the present experiment, dissonance may have been created in the following ways. First, the person may be aware that he could have taken a position other than the one he has assumed; second, he may feel that other people will not agree with his stand. There is evidence (Festinger, Riecken, & Schachter, 1956) that dissonance arising from these sources can produce affiliative behavior.

A Contradiction (?)

The hypothesis that unchanging (certain) deviates in the Peer condition would want to join a discussion group composed of other people from their psychology classes contains a possible logical contradiction if it is presumed that the subjects were perceptive and logical. The distribution of opinions that these subjects, Peer deviates, had seen told them that nearly 90% of the potential discussants would disagree with them. Festinger (1954) states that for social comparison to produce stable evaluation of opinions, the opinions of others must be close to one's own. Clearly there would be few if any people in a discussion group composed of six to eight people with opinions close to those of the deviate, given the distribution shown. Why, then, did the subject's behavior conform to the predictions of this apparently contradictory hypothesis?

Several explanations are possible. One is that the subjects may simply not have made the logical inference necessary for them to have predicted the probable composition of the discussion groups. At the other extreme of the subjects' perceptive ability, it would probably be implausible to assume that subjects would hypothesize that groups would be composed largely of deviates; that subjects would realize that conformers would not want to join discussion groups because their opinions had already been evaluated.

A theoretically more interesting and experimentally testable hypothesis is that unchanging deviates considered but disregarded the probable composition of the discussion groups. They may have had such strong needs for more evaluative information that a potential comparison even with quite divergent views was sufficiently attractive to arouse affiliative tendencies. According to this reasoning, it would be expected that, given a choice, unchanging deviates would rather join a discussion group composed of people whose opinions were similar to their own.

Birth Order

As mentioned in the introduction, the hypothesis concerning the effects of ordinal position stemmed from Schachter's (1959) experiments on anxiety and affiliation. He invoked dependency as a construct to account for the greater affiliative tendencies of early borns under conditions of high anxiety.

With reference to birth order and dependency, the present study fills in the missing cell in a 2×2 table summarizing the results of four experiments in which emotions and opinions were in need of evaluation, and changes as a result of social comparison and affiliative tendencies were measured. Compared with later borns: early born subjects showed greater changes in (*a*) emotional reactions (Wrightsman, 1959) and (*b*) opinions (Ehrlich, 1958) as a result of social comparison; early borns also manifested greater affiliative tendencies when in need of (*a*) emotional comparison (Schachter, 1959) and (*b*) opinion comparison (present study). The complementary results of these four studies strongly suggest that the static variable of birth order is related to the more dynamic concept of dependency.

SUMMARY

The hypothesis that self-evaluative needs concerning opinions are an important determinant of group formation was tested experimentally.

Subjects were presented with a problem and asked to state their opinions. Adequacy of opinion evaluation was varied at four levels along a continuum by telling subjects in different conditions the opinions of various groups of other people. Affiliation tendencies

were measured by assessing the level of interest the subjects had in joining discussion groups to talk about the opinion in need of evaluation.

Results gave strong support to the hypothesis. Affiliation tendencies varied according to the strength of evaluative needs.

Consonant with Schachter's (1959) findings regarding emotional evaluation and affiliation, the affiliative responses of first born and only subjects were more strongly effected by varying evaluative needs than were those of the later borns.

REFERENCES

BACK, K. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

EHRLICH, D. Determinants of verbal commonality and influencibility. Unpublished doctoral dissertation, University of Minnesota, 1958.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., GERARD, H., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327–346.

FESTINGER, L., RIECKEN, H., & SCHACHTER, S. *When prophecy fails.* Minneapolis: Univer. Minnesota Press, 1956.

FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.

GERARD, H. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249–272.

HOCHBAUM, G. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *Amer. sociol. Rev.*, 1954, **6**, 678–687.

RADLOFF, R. Opinion and affiliation. Unpublished doctoral dissertation, University of Minnesota, 1959.

SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford Univer. Press, 1959.

WRIGHTSMAN, L. S. The effects of small-group membership on level of concern. Unpublished doctoral dissertation, University of Minnesota, 1959.

(Received May 13, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 586-592

FEAR AND SOCIAL COMPARISON

HAROLD B. GERARD
Bell Telephone Laboratories, Murray Hill, New Jersey
AND JACOB M. RABBIE[1]
Yale University

In a series of experiments, Schachter (1959) has studied some aspects of affiliation by inquiring into the instrumentalities served by being with others when a person is exposed to a stressful situation. The subject in his experiments anticipated either a weak or a strong shock and then indicated his desire to wait either with others or alone before getting shocked. Schachter suggests five plausible "explanatory possibilities" that might each account for the experimentally demonstrated positive relationship between fear and affiliation. Being together with others might provide a person with: a collective means for avoiding the experiment, i.e., escape; the attainment of cognitive clarity regarding the impending danger; comfort and support; distraction; or a means for attaining clarity regarding his "worked up" body state, i.e., self-evaluation. By varying aspects of the projected waiting situation described to the subject, and on the basis of internal analyses of his data, Schachter rules out all but the comfort and self-evaluation hypotheses and proceeds to lay heavier emphasis upon self-evaluation. The present experiment was prompted by the following difficulties inherent in Schachter's method of inferring such a desire for self-evaluation via social comparison: the unduly circumscribed list of "plausible explanatory possibilities," the reliance upon internal correlational analysis to eliminate the three other possibilities, and the possible presence of experimental constraints limiting the operation of other motives which may, under ordinary circumstances, produce affiliation.

Let us accept the basic assumption underlying comparison theory, namely, that the person will attempt to reduce cognitive unclarity regarding some aspect of himself. If information from others is available to him regarding that aspect of *themselves*, he should tend to employ the information in attempting to evaluate *himself* in that respect. A number of kinds of internal cognitive unclarity regarding an aroused state may exist: The individual may want to *label* the emotion, *measure* its intensity, determine whether, in a normative sense, it is *appropriate*, or discover whether or not he is *reacting adaptively* to the impending situation. For the present, we single out the aspect of intensity. We assume, in line with comparison theory, that intensity information from others provides the individual with a frame of reference within which to evaluate the intensity of his own reaction.

Schachter (1959) created his high and low fear treatments by verbal instructions while the subjects were seated in full view of each other. It would seem, however, that inducing fear in a face-to-face setting probably does not control for a considerable amount of communication, as some of Schachter's own research indicates. We therefore decided to run our experiment under restricted conditions, the subjects receiving their instructions in isolation.

If a person is uncertain as to the intensity of his emotional reaction, he should seek information from others that will help him measure it. To the extent that he has information from others regarding the intensity of their reactions, his desire to be with them prior to the impending experience should be reduced.

METHOD

Two levels of fear were induced by telling the subject that he would receive either a strong or a weak electrical shock. He was led to believe that, with the aid of certain electrodes attached to him, we could accurately determine the extent of his emotional reactions. Within the two fear treatments three information treatments were administered. In one treatment the subject received information as to the degree of his own reaction and the reactions of three other subjects who were seated in adjacent cubicles. In a second treatment the subject received information as to his reaction only. In the third treatment no information was given although the same electrodes were in place. This last treatment was in all essential respects a replication of

[1] We wish to thank George S. Rotter for help in running the experiment and Stanley Schachter and Arthur R. Cohen for helpful, critical comments.

Schachter's basic experiment with the exception that subjects were isolated from one another during the fear induction and four subjects were used per group wherever possible (Schachter used from five to eight subjects).

The subject was asked to state whether he wanted to wait alone or together before getting shocked and to what degree he desired the waiting situation he chose. Questionnaire and other data were then obtained.

Subjects

Twenty-nine male and 71 female undergraduates at Brooklyn College served as paid volunteers. The sex distribution was determined by the number of available volunteers. Each group consisted of either all male or all female subjects.

General Setting

Subjects were seated in adjoining cubicles. A pair of Scherping finger electrodes were attached to the index and ring fingers of the subject's nonpreferred hand. In addition, an EKG electrode was attached to his forearm, which was immobilized by strapping it into a specially designed arm rest. The subject was informed that the finger electrodes registered skin resistance and that the arm electrode measured muscle tremor. By employing this information, he was informed, we could accurately determine his degree of emotionality. We were interested, the experimenter continued, in determining his emotional reaction to various kinds of situations that he would soon experience. In actuality, the forearm electrode was a dummy. The finger electrodes, however, were real and were connected to one of the inputs of a four-channel constant current galvanometer circuit designed to measure skin resistance.[2] Since changes in skin resistance are known to reflect stress, we were provided with a physiological check on the fear manipulation. All of the instructions for each of the six experimental conditions were prerecorded on magnetic tape. Constant within-condition exposure was thereby achieved. Instructions were received by the subject over a headset with a rubber cup over each earpiece that fitted tightly against his head. This arrangement achieved some degree of auditory isolation.

Manipulation of Fear

Upon entering the laboratory, the subject saw a rather impressive array of electrical equipment with many switches, variously colored lights, and many lengths of wire and cabling. This was actually the control and recording equipment used in the experiment but could easily be viewed in retrospect by the subject as equipment that could be used to deliver electrical shocks. In the high fear treatment a bold sign saying "Caution, High Voltage" was prominently on display. The two experimenters were dressed in white laboratory coats and introduced themselves to the subjects. In the High Fear treatment one of the experimenters had a stethescope bulging out of his pocket and a sphygmomanometer was conspicuously on display. Both these features were absent in the Low Fear treatment. A short waiting period ensued which was used to establish the subject's base level skin conductance. In both treatments the subject was told that the experiment was concerned with the effects of electrical shock.

In the High Fear treatment the instructions continued:

> We would like to give each of you a series of electrical shocks. Now, I feel that I must be completely honest with you and tell you exactly what you are in for. These shocks will hurt; they will be somewhat painful. As you can guess, if in research of this sort we are to learn anything at all that will really help humanity, it is necessary that our shocks be intense. What we will do is hook you up to the apparatus, give you a series of electrical shocks and then determine your reaction to the shocks. Again, I do want to be honest with you and tell you that these shocks will be somewhat painful, but of course they will do no permanent damage.

In the Low Fear treatment the following instructions were substituted for the foregoing ones:

> We would like to give each of you a series of very mild shocks. I assure you that what you feel will not in any way be painful. It will resemble more a trickle or a tingle than anything unpleasant. We will put electrodes on your hands, give you a series of very mild shocks and then determine your reaction to the shocks. I assure you there is absolutely nothing to be worried about.

At the beginning of the Low Fear instruction the subjects were told that "they would enjoy the experiment." This remark was omitted in the High Fear treatment. In all conditions it was emphasized that the electrodes which were attached to the person were for recording purposes only and would not be used for the administration of the shocks.

Manipulation of Information

After the fear induction we established three information treatments: the subject received his emotionality index and the indexes of the others (Inf-S + O), only his own emotionality index (Inf-S), no information (No-Inf).

The Inf-S + O treatment was introduced as follows:

> As you no doubt may guess, even the anticipation of getting a shock may affect you emotionally and hence produce physiological reactions. In front of you there are four microammeters which have been calibrated to register the emotionality index I referred to earlier. We thought you would like to see how you yourself and the others are reacting to the prospect of getting shocked. As you can see each meter goes from 0 to a maximum of 100. The first meter to register will show your own index. Then the other meters will be connected so that you may see how emotional the others are.

At this point the subject's meter was switched on and a few seconds later the other three meters were energized. Each subject was in exactly the same parallel situation. The third meter from the left was the first to read, this presumably being the subject's own emotionality index. Preset potentiometers in the meter circuits caused the meters to read 79, 80, 82, and 81

[2] The instrument was designed by J. Kouzoujian.

in the High Fear condition and 29, 30, 32, and 31 in the Low Fear condition. The third figure was presumably the subject's own index. In both treatments we arbitrarily set the subject's reading slightly higher than the rest. On an a priori basis we believed that either a lower reading or one which was at the average would tend to reduce the subject's actual emotional reaction and thus confound the fear manipulation. Since we know absolutely nothing, as yet, about the disturbing or reassuring effects of this kind of information, the choice seemed reasonable. In the Inf-S treatment, the subject was confronted by a single meter which registered 82 in the High Fear and 32 in the Low Fear treatment.

After the information manipulations were administered the subject was told that there would be a 10- or 15-minute wait while we prepared the equipment and while he awaited his turn to be shocked. He was given a choice of waiting together with the others in a room nearby or alone in a separate room. He was told that, on the basis of the choices made, we would make the most appropriate waiting arrangement. He depressed one of two buttons to indicate his choice. He was then asked to indicate the strength of his desire either to wait alone or together (depending upon which he had chosen). He did this by setting a dial in the center of his console at a position between 1 and 11. His dial varied the setting of a potentiometer which caused a milliameter, which was calibrated from 0 to 100, to register differentially. We now had the basic information which we could use to test the comparison hypothesis.

Check on the Fear Manipulation

In addition to the skin resistance record, other checks on the success of the fear manipulation were included.

After the questionnaire was completed the subject was asked, in the two information treatments, to answer the following question printed on a separate sheet of paper:

> On occasion we have found the individual's actual feelings, that is, his subjective experience, does not exactly correspond to his physiological emotionality index. If you had been asked to guess your meter reading before seeing it, assuming 100 to be maximally emotional, what would you have guessed?

A blank was left for the subject to indicate his guess.

The final event in the experiment was a further check on the fear induction. The subject was given an opportunity to change his mind as to whether or not he wished to participate. If he chose not to participate, however, he would forfeit the honorarium of $1.50 which he was promised. After the subject made this choice, the experiment was over.

The experimenter then revealed the deception and described, in brief, the purpose of the experiment. The subject was then asked not to reveal anything about the experiment to those who had not as yet participated.

RESULTS AND DISCUSSION

Effectiveness of the Fear Manipulation

Three independent checks of the fear induction were made: change in skin conductance, the subject's guess (in the information treatments) as to what his meter reading would be, and the subject's decision to remain or leave prior to the supposed shock administration.

We subtracted the subject's base conductance immediately before the fear induction from his level at the end of the induction. The average increase in conductance was greater in the High Fear than in the Low Fear treatment ($p < .001$ by t).

The mean for the guess as to the meter reading was 60.00 for the High Fear and 41.14 for the Low Fear treatment ($p < .001$ by t).

The vast majority of subjects chose to go through with the experiment. However, in the High Fear treatment 11 subjects chose to leave as compared with only 1 in the Low Fear treatment ($p < .01$ by chi square).

These measures leave little doubt as to the overall success of the fear induction.

Fear, Information, and Affiliation

We first turn to an examination of the data relevant to a test of the comparison hypothesis. Only 12 of our 100 subjects chose to be alone, with no systematic relationship to our treatments. With regard to the choice of "together" or "alone," there is a basic procedural difference between this and Schachter's experiments. When asked to choose "together" or "alone" Schachter (1959) also provided the subject with a "don't care" alternative. In his analysis he combined the "alones" and "don't cares" and contrasted them with the "togethers." The percentage of subjects choosing "alone" in our No-Inf treatment is the same as in Schachter's first experiment (approximately 9% with no difference between High and Low Fear). His results are based principally upon a comparison between those subjects choosing "together" and those choosing "don't care." The small proportion of "alones" in both this and Schachter's experiment suggests that these individuals are probably deviant in some related respects. We shall return later to a consideration of those subjects who chose "alone." The omission of the "don't care" alternative, thereby forcing a choice between "together" and "alone," makes our dependent measure not strictly comparable with Schachter's. In discussing the strength of affiliation we will re-

TABLE 1
FEAR, INFORMATION, AND STRENGTH OF AFFILIATION FOR SUBJECTS CHOOSING "TOGETHER"

	No-Inf	Inf-S	Inf-S + O
High Fear	66.80 (15)	70.50 (16)	55.09 (11)
Low Fear	54.53 (17)	64.12 (17)	47.67 (12)

fer to the data for those 88 subjects who chose "together."

Table 1 presents the mean desire to be together for all six treatments. A difference between fear treatments is indicated ($p < .10$ by a one-tailed t test; using a median split, $p < .05$ by chi square). A test of the comparison hypothesis involves evaluating the difference between the two information treatments. This difference is in the predicted direction ($p = .05$ by t) thus supporting the hypothesis. The data in the No-Inf treatment complicate the picture in that a quadratic trend is suggested with increasing information ($p < .07$ by t), i.e., both the No-Inf and Inf-S + O treatments are low relative to the Inf-S treatment. Given that the individual was uncertain as to the extent of his emotion, providing him with information about himself would serve to reduce this uncertainty and hence reduce his desire for comparison. The addition of information about the others would be expected to reduce his comparison desire even further. How can we then account for the relatively high level of affiliation engendered in the Inf-S treatment?

Other data, which are at hand, might shed some light on the matter. We determined, for each subject in the Inf-S and Inf-S + O treatments, the difference between his meter reading (82 in the High Fear and 32 in Low Fear treatment) and what he later guessed it would be. We correlated the absolute value of this difference with his desire to be together. The coefficients are: High Fear Inf-S, .22; Low Fear Inf-S, .72; High Fear Inf-S + O, −.03; Low Fear S + O, .15. The Pearson coefficient for the Inf-S treatment is higher than for Inf-S + O ($p < .02$ by t). When the two sources of information (his subjective experience and his meter reading) about the subject's reaction were discrepant, it seems likely that ambiguity was induced which, in turn, engendered comparison motivation. Given that the information was already available (in the Inf-S + O treatment), the ambiguity was resolved. The correlation is quite high, interestingly, within the Low Fear Inf-S condtition whereas it is not significantly different from zero in the High Fear Inf-S treatment. In both fear treatments discrepancy is, to a great extent, a function of guessed emotionality. In the Low Fear treatment a large discrepancy occurred when the subject guessed that his meter reading would have been considerably *higher* than 32, whereas in the High Fear treatment a large discrepancy occurred when the subject guessed that his meter reading would have been considerably *lower* than 82. In the Low Fear treatment, therefore, guessed emotionality and discrepancy act in concert, whereas in the High Fear treatment they act in opposition.

On the basis of this possible confounding of discrepancy with emotionality, a higher correlation under Low Fear than under High Fear is to be expected. This difference which is evident under Inf-S does, however, not occur under Inf-S + O. That this is not so in the latter treatment suggests a reduction of comparison motivation. Since a covariance analysis of the present data is not appropriate, additional research is required in which discrepancy and the individual's absolute phenomenological level of fear are varied independently.

One may tentatively suggest, on the basis of these data, that the Inf-S treatment may have increased the salience of the individual's emotional state by confronting him with a discrepancy. It is also possible that the mere fact of knowing the absolute value, so to speak, of a measurement of some aspect of the self induces a desire for knowledge of relative standing. Results in a recent experiment by Radloff (1959) suggest an analogous process in the use of opinion comparison. When our subject found confirmation for his subjective feelings, i.e., where there was agreement between the meter reading and what he would have guessed it to be, he had little need for additional information. However, where the discrepancy was large his curiosity was aroused as to how the others were reacting. The self-information treatment may have induced ambiguity where none or very little existed in the first place. Given a discrepancy, however, it is not clear just what kind of information (assuming the subject was seeking some) he wanted. Perhaps he wanted to reach some compromise, i.e., to use the information others might provide as a

means for reconciling the two conflicting sources of information with which he was faced. In so doing he might assign different weights to the two sources and then proceed to average them in some way in order to reach a compromise. Another motivation for affiliating may have been to determine if he was interpreting the scale correctly. Perhaps the emotional levels he coordinated to the end points were somehow wrong. Still another variant of comparison motivation here might be the individual's desire to determine if some constant discrepancy was experienced, i.e., perhaps the others' subjective feelings also differed by an amount similar to the discrepancy with which he was confronted. These represent some of the possibilities which should be explored.

Two additional difficulties, one methodological and one statistical, occur in interpreting the correlation between the discrepancy of subjective feelings from the meter reading, and affiliation. The emotionality guess was a retrospective view by the subject of himself and is therefore subject to a number of sources of contamination. Also the variance of guesses within the Inf-S + O treatment tends to be smaller than the variance within the Inf-S treatment ($p < .10$ by F). This homogenization would follow since additional reference information had been provided and all the readings were close to one another. The chances for a correlation in the Inf-S + O treatment would be reduced relative to the Inf-S treatment by allowing individual error a wider relative latitude to affect the correlation. In order to explore the problem further, this feature of the situation would have to be controlled.

Subjects Who Chose "Alone"

Of the 6 subjects in the High Fear information treatments (Inf-S and Inf-S + O combined) who chose "alone," 5 had emotionality index guesses of 70 or higher. Of the 27 subjects who chose "together" in these treatments only 12 had guesses of 70 or higher. The data on "alone" choices are sparse. However, the interaction just noted suggests that under certain circumstances, which we are unable as yet to specify, an emotionally aroused individual prefers seclusion to the company of others who are faced with the same plight. Another line of evidence suggests a positive relationship between fear and preference for isolation. For each of the eight subjects in the information treatments who chose "alone" we compared his guessed emotionality with the strength of his desire to be alone and found a Spearman coefficient of .76, which for an N of eight, is significant at the .02 level. Thus high emotionality for "alone" subjects relative to those who chose "together" and the correlational evidence just discussed, suggest that an "alone" choice did not reflect indifference toward being together as Schachter's (1959) grouping of the data might lead us to suspect. Sarnoff and Zimbardo (unpublished) report that the induction of oral anxiety (where a psychosexual conflict is aroused) predisposes the subject toward being alone rather than together, the reverse being true for fear (where arousal does not involve a psychosexual conflict). The latter is comparable to our manipulation. Those few subjects in our experiment who chose to be alone in the High Fear treatment may thus have suffered a diffuse anxiety reaction of the kind generated by Sarnoff and Zimbardo in their anxiety manipulation. Also, we have noted that the preponderance of subjects who chose "alone" were men and that their emotionality was high. A public display of emotion is, in our culture, considered unmanly. Experiencing the emotion in private obviates the necessity of hiding it from others and removes the risk of having it displayed inadvertently. This suggestion brings into focus the normative aspect (the reaction others would presumably approve) of emotional comparison which has thus far received little attention.

Sex, Birth Order, and Affiliation

Schachter (1959) found a positive relationship between a questionnaire item measure of anxiety and the tendency to affiliate. We decided to examine, irrespective of experimental treatment, the relationship between emotionality as measured by an increase in skin conductance and desire to be together. Here we find an interesting male-female differential. Using median splits for both variables we find a positive relationship for women, namely, that a large increase in skin conductance during the fear induction is

associated with a strong desire to be together, whereas a small increase in skin conductance is associated with a weak desire to be together ($p < .02$ chi square). For men an opposite trend occurs ($p < .10$ by chi square). The considerably smaller N for men makes statistical comparison unstable. This difference in trend is quite interesting and brings us to a closer examination of the effect of the sex variable and of the birth order effect reported by Schachter.

On the questionnaire we asked the subject to indicate the age and sex of siblings, if any, and to indicate his own age. The subjects in Schachter's (1959) experiments were all University of Minnesota coeds whereas we used both men and women. We find a difference in accordance with Schachter's finding, i.e., using a median split in classifying for high and low affiliation we find that 42% (21/50) of first born and only children have low affiliation scores whereas the percentage for later borns is 61 (23/38, $p < .05$ by chi square). If we examine men and women separately we find quite a striking contrast in the relationship. Thirty-four percent (14/41) of first born and only women had low scores whereas among the later born women the percentage is 71 (20/25, $p < .01$ by chi square). For men the relationship appears reversed. Of the 29 men, 10 chose "alone" (only 2 of 71 women chose "alone"). Among the 19 men who chose "together," 77% (7/9) of the first born and only children and 30% (3/10) of later borns had low scores ($p < .07$ by an exact test). The mean affiliation scores, combining across fear treatments, are: first born and only men, 45.1; first born and only women 67.0; later born men, 61.4; later born women, 55.7. A large difference occurs between men and women first borns, whereas what appears as a reversal for later borns is not a significant difference. Of the 10 men who chose "alone," the 4 first borns had an average desire to be alone of 52.7 whereas the average desire of the 6 later borns was 44.3.

Schachter (1959) finds some evidence for a relationship between birth order and emotional reactions to stress, namely, that first born and only women tend to react more strongly than later borns. We examined, for all subjects, the relationship between birth order and change in skin conductance during the fear manipulation. These data are presented in Table 2. We observe the High versus Low Fear difference referred to earlier as well as sex and birth order effects. In the High Fear treatment we see an interaction between sex and birth order which parallels the affiliation data ($p < .05$ by t). The relationship between birth order and affiliation might therefore be attributed to a difference in emotional reaction to the fear induction and not to a difference in the reaction which occurs given emotional arousal.

TABLE 2

BIRTH ORDER, SEX, FEAR TREATMENT AND CONDUCTANCE CHANGE (micromhos)[a] DURING THE FEAR INDUCTION

	High Fear		Low Fear	
	Men	Women	Men	Women
First Born	1.67(6)	2.58(20)	0.00(7)	0.67(21)
Later Born	3.74(5)	1.31(14)	0.79(11)	0.30(12)

[a] Skin conductance data for four women are missing due to an equipment failure.

Pursuing the problem further becomes difficult due to the rapidly shrinking cell size with further breakdowns. Examining the data for men becomes impossible for this reason so we shall confine the immediately following analysis and remarks to the data for women. In order to determine if the relationship between sex, birth order, and affiliation was to be accounted for solely by the difference in emotionality just observed, we attempted to control for emotionality. In order to have a sufficiently large sample to examine, we combined the fear treatments. Before doing so, we dichotomized both the affiliation and conductance variables using a median split within fear treatments and then examined the resulting contingency for first born and only versus later born women. This within-treatment split corrects for the differential effect of the fear treatments upon conductance change and affiliation. We find that a conductance differential for first born and only women is correlated with affiliation ($p < .02$ by chi square), whereas the trend for later borns is not significant. This evidence leads us to suspect that the relationship between birth order and affiliation is due, in part at least, to

a difference in reaction to fear. This corroborates Schachter's finding.

In the light of Schachter's (1959) conclusion concerning the relative overprotection of first born children, and the subsequent development of "greater dependence in first born than in later born children," how can we account for the present difference in reaction between first born and only women and their male counterparts? One possible interpretation involves not so much the attitude learned by the child due to the strength of the succorance relationship with his mother as those that she communicates. The parent conveys society's norms to the child. One set of norms that is communicated is how a male or a female should behave. In interacting with her male child the mother treats him as a male, rewarding male-like behavior to the exclusion of female-like behavior. The reverse is true for females. The behavior of first born and only children in our experiment exemplifies the Victorian norms of the stalwart man and dependent woman which still probably characterize the ideal man and woman in contemporary western culture. What about the behavior of later borns? A distinguishing feature between first and later born children is that, during the formative years (months?) the first born child has contact only with his or her parents whereas the later born child is exposed both to parents and to one or more older siblings. An older sibling will not normally display different behavior toward the younger child because of that child's sex; he or she will treat the other as just another child. At some age an individual, after fully incorporating society's sex norms, may assume the responsibility, so to speak, of communicating them to younger siblings. Our data in the present experiment are unfortunately too sparse to study the effects upon affiliation of the age differential between a child and his next older sibling.

SUMMARY

The present experiment was stimulated by some work by Schachter (1959) in which he applied Festinger's (1954) theory of social comparison to account for the arousal of affiliative tendencies when the individual is exposed to a threat. We attempted to test directly the tenability of explaining the relationship between fear and affiliations by social comparison theory. According to the theory, the individual is motivated to compare with others some aspect of himself, e.g., an opinion he holds, an ability he possesses, or an emotional experience whenever other means of achieving certainty about that aspect are lacking. Using a modified version of Schachter's fear arousing situation, in which electric shock was threatened, we provided the individual with information as to the intensity of emotion experienced by three other individuals seated in adjacent cubicles. According to the theory this should reduce the desire for affiliation. Support was found for the theory. Other motives also appear to have been operative as well as strong sex differences in reaction to threat. The effect of sex appears to be linked with the order of birth in the family.

REFERENCES

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

RADLOFF, R. W. *Opinion and affiliation.* Unpublished doctoral dissertation, University of Minnesota, 1959.

SCHACHTER, S. *The psychology of affiliation.* Minneapolis: Univer. Minnesota Press, 1959.

(Received May 24, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 593-597

A COMPARISON OF GENERALIZATION AND ADAPTATION-LEVEL AS THEORIES OF CONNOTATION

JEROME E. PODELL
San Francisco State College

To *connote* commonly means "to suggest . . . in addition to the explicit or denoted meaning: as, the word *mother* means (denotes) 'female parent,' but it generally *connotes* love, care . . ." (*Webster's New World Dictionary of the American Language*, 1951). In other words, the connotation of a word may be expressed qualitatively by another word. Recently, however, Osgood, and his associates (Jenkins, Russell, & Suci, 1958; Osgood, Suci, & Tannenbaum, 1957) have shown that connotation may be expressed quantitatively as values on certain dimensions of meaning. In addition, these investigators have shown that the major dimensions of connotation, are, in order of importance, evaluation (a good-bad continuum), strength, and activity. Since the major portion of variance in ratings on connotational scales has been accounted for by these dimensions, it seemed quite worthwhile to examine in detail the relationship between verbal (qualitative) and quantitative connotation.

Osgood et al. (1957) have explained the connotation process by means of a mediated generalization model. This view suggests the following relationship between verbal and quantitative connotation, viz., if two words have the same quantitative connotations, i.e., equivalent sets of values on the dimensions of the semantic space, then each word should be a verbal connotation of the other. More concretely, when a person is set to respond by giving a verbal connotation, the most likely response is that word having quantitative connotation most like that of the stimulus word. This statement summarizes what will be hereafter referred to as the generalization view. Although generalization is a well-substantiated theory in the field of learning, its application to the field of meaning, as Osgood et al. (1957) point out, demands further experimental investigation. Furthermore, one may accept the dimensional approach to connotation without being bound logically to accept the generalization view.

A likely alternative to this view is the adaptation-level theory of Helson (1947). This theory, while developed largely within the context of perception, has been shown to apply also to the field of meaning (Helson, Dworkin, & Michels, 1956). It differs from the generalization view in that it takes account of the quantitative character not only of the stimulus, but also of the population of possible responses. The present study aimed to discover which of these two theories might yield the more accurate deduction relative to verbal connotation of words referring to personal traits.

Studies by Asch (1946) and Kelley (1950), prototypes of the experiment reported here, may be viewed as experiments in verbal connotation, if the following assumption be granted. Since responses in these studies were rarely synonymous to the stimulus words, impressions of persons formed from trait-terms involve connotation. Each subject in the Asch (1946) and Kelley (1950) experiments was given one of two short initial descriptions of persons, each consisting of identical lists of traits except for the inclusion of the word COLD in one list, and the word WARM in the other. Subjects formed impressions on the basis of these initial lists, and subsequently selected additional traits appropriate to the person described earlier. It was found that impressions formed under the "warm" condition differed markedly from those formed under the "cold" condition, and the theory was advanced that the differential effects of the traits WARM and COLD were attributable to their being "central" traits (Asch, 1946).

A general explanation of the results obtained by Asch (1946) and Kelley (1950) can be provided by invoking either the mediated generalization theory (Osgood et al., 1957) or adaptation-level theory (Helson, 1947). According to the generalization view, the process by which the initial impression influences subsequently chosen traits depends upon mediated generalization along various dimensions. Since this view implies that the response

tendency is a direct function of stimulus response similarity, it follows that the probability that a trait will be chosen is a direct function of the similarity between the trait and the initial impression with respect to relevant dimensions of meaning. Furthermore, if the evaluative dimension (favorableness) was, as usual, the dominant determinant of connotation, we can simplify our discussion by inferring that the traits chosen would be similar in favorableness to the traits initially presented. If we assume that warm and cold are good and bad, respectively, then the warm and cold impressions would mediate responses similar to themselves with respect to favorableness. This interpretation of the results obtained by Asch (1946) and Kelley (1950) is based on the hypothesis that stimulus words and response words in this situation should have identical values on the relevant dimension of meaning.

Adaptation-level theory implies additionally that the subsequently chosen traits should be affected not only by the favorableness of the initial impression, as above, but also by the favorableness of the whole set of potential responses. In adaptation-level terms, the initial list functions as a background (frame of reference), while the check list functions as the comparison series stimuli; each list having a net stimulus value equal to the average of the values of its component parts. According to this view, the average favorableness of the traits chosen should fall somewhere between that of the initial list and that of the check list. Accordingly, the present study deals with evaluative connotation as affected by the net favorableness of the initial list of traits (background), and the net favorableness of the check list of traits from which the subjects subsequently select their responses (series stimuli).

METHOD

Construction of the Favorableness Scale

Seventy-eight adjectives were scaled by a method similar to that of Mosier (1941). Each word was typed on a 3″ × 5″ card, and judges were required to sort them according to a seven-category scale of favorableness of personal traits; 1 being the most favorable, and 7 being the least favorable. No restriction was placed on the number of traits assigned to each category. Twelve undergraduate students served as judges, six from the University of California at Berkeley, and six from San Francisco State College. Half the judges from each institution were men, and the other half, women. The mean scale value was computed for each word, and then a scale was constructed by selecting, as far as possible, one word for every two-tenths of a scale interval. The scale is shown in Table 1. Three additional words near the ends of the scale (indicated by asterisks) formed the Initial Lists (background stimuli) described below.

Experiment

Subjects. Forty-eight students in the General Education Psychology course at San Francisco State College served as subjects.

Procedure. All subjects were given an Initial List of three trait-terms, and requested to form therefrom an impression of a fictitious person. They then were presented with the Check List (Table 1 minus the asterisked terms), and instructed to check four terms that would best describe the person. There were two independent groups of subjects; one group given the Favorable Initial List consisting of the words CAPABLE, INFORMED, and CONSIDERATE, and the other group given the Unfavorable Initial List consisting of the words MEDIOCRE, SMUG, and HARSH. These six words and their scale values are indicated by asterisks in Table 1. The mean favorableness values of the three lists were as follows: Favorable Initial List, 1.8; Unfavorable Initial List, 5.3; and Response Check List, 3.6. Word order for all lists was counterbalanced by presenting each Initial List in the above order and its reverse, and by presenting the Check List in a random order and its reverse. The four resulting Initial Lists and the two Check Lists were administered in all of the eight possible combinations, with six subjects receiving each combination. All subjects were given a two-page pamphlet which bore on its first page the following instructions:

> This is a test of your ability to judge personality. Suppose that all you knew about a person is that people say he is capable, informed, and considerate (mediocre, smug, and harsh). Now turn the page. (The second page bore the following.) Now imagine what sort of person this must be. Look at this list of words. Then select four words from the list which strike you as most likely to be true of the person. Here is the list. Check four words which you feel would be true of the person described on the first page. (The Check List followed.)

RESULTS

The median scale value of the four checked words was employed as the score for each individual since the response distributions were skewed. The mean median of the scores obtained by the group receiving the Favorable Initial List was 1.9, while the mean median obtained by the group receiving the Unfavorable Initial List was 5.1. The two distributions showed no overlap, and the probability of this outcome under the null hypothesis is less than 0.1%.

Under the Favorable condition, the dif-

TABLE 1
TRAIT-TERM LIST

Trait	Scale Value	Trait	Scale Value
Genuine	1.2	Quiet	3.8
Self-controlled	1.4	Rural	4.0
Efficient	1.5	Crafty	4.2
Capable*	1.6	Abrupt	4.5
Kind	1.8	Ordinary	4.7
Informed*	1.8	Erratic	4.9
Considerate*	2.0	Domineering	5.0
Adaptive	2.0	Mediocre*	5.0
Warm	2.2	Cold	5.1
Receptive	2.4	Indifferent	5.3
Intent	2.6	Smug*	5.3
Zealous	3.0	Harsh*	5.7
Forward	3.1	Inconsistent	5.7
Proper	3.2	Sour	5.9
Robust	3.4	Offensive	6.6
Reserved	3.6		

Note.—Asterisks indicate three additional words near the ends of the scale forming the Initial Lists (background stimuli).

ference of 0.1 between the mean median response and the mean of the Initial List, 1.8, was not significant. However, under the Unfavorable condition, the difference of −0.2 between the mean median response and the mean of the Initial List, 5.3, was significant beyond the 1% level (Wilcoxon, 1949).

Since under both the Favorable and the Unfavorable conditions the central tendency of response fell between the means of the Initial and Check Lists, the data were further analyzed according to adaptation-level theory by computing the relative weights of the Initial and Check Lists as determiners of response. This was done by means of a modification of the formula developed by Helson (1947) for lifted weights. This formula (Formula 1) states that the log of the weight of the stimulus judged as neutral (neither heavy nor light), plus a correction for the stimulus interval, is equal to the weighted mean of the mean log of the series stimuli and the log of the comparison stimulus, with the series stimuli being weighted three times as much as the comparison stimulus.

$$\log (\mathrm{AL} + 0.75\mathrm{d}) = [(3\Sigma \log X_i/n) + \log \mathrm{C}]/4 \quad [1]$$

This formula was modified to suit the present study as follows. Logarithms were eliminated because the words had values on a psychological rather than a physical scale. Hence there was no need to employ the Weber-Fechner psychophysical relationship. The triple weight for the series (X_i), and the unit weight for the comparison stimulus (C) were changed from constants to variables, since the calculation of these weights was a problem for the present study. In addition, the stimulus-interval factor (0.75d) was ignored as too refined for an exploratory study. Finally, the sum of the weights (w) was made equal to 1. The results of these modifications are shown in Formulae 2 and 3. It may be noted that Formula 2 is identical in form to that used in physics for the equilibrium of parallel forces (Sears, 1946).

$$\mathrm{AL} = [(\Sigma X_i/n_i)w_i + (\Sigma X_j/n_j)w_j]/w_i + w_j \quad [2]$$

$$w_i + w_j = 1 \quad [3]$$

In these equations, X_i and X_j refer to favorableness values of traits in the Initial and Check Lists, respectively; n_i and n_j refer to the number of traits in each list; and w_i and w_j refer to the weights of the two lists. The adaptation-level was taken as the mean median response under the two experimental conditions, i.e., 1.9 for the Favorable condition and 5.1 for the Unfavorable condition.

The weight of the Check List under the Favorable condition was 0.05 while its weight under the Unfavorable condition was 0.16. The corresponding weights for the Initial Lists were 0.95 and 0.84. Thus the weights of the Initial Lists in determining the adaptation-level ranged from 5 to 19 times as great as that of the Check Lists, or conversely, the weight of the Check List in determining the adaptation-level ranged from $\frac{1}{5}$ to $\frac{1}{19}$ that of the Initial List.

DISCUSSION

Dimension Theory of Meaning

The present results provide support for the dimension theory of meaning. Specifically, it was found that the favorableness of words chosen to match an initial list was strongly and reliably affected by their favorableness. Despite the fact that the traits were not controlled for "loading" on other dimensions such as *activity* and *potency* the favorableness dimension alone had a very substantial effect in determining judgments. This result is

consistent with the factor analytically derived primacy of the favorableness dimension, and the theory that verbal and quantitative connotation are related.

Certain other studies which also concern dimensions of meaning differ from the present one in that the responses were glandular secretions (Razran, 1935; Riess, 1940) or motor responses (Osgood, Suci, & Tannenbaum, 1957, pp. 155–160) rather than words. In contrast to these studies, the present one concerns symbolic responses to symbols rather than sign responses to symbols.

Theories of Adaptation-Level and Generalization

One may logically adopt the dimension theory of meaning without necessarily accepting the theory of mediated generalization; adaptation-level providing a view alternative to the latter. With respect to the present experiment, the difference in implications of the two theories concerns the factors that determine the central tendency of the responses. While the generalization theory implies that the response depends only on the stimulus, the adaptation-level theory implies that the response depends both on the initial stimulus and the set of possible responses. Specifically, the latter theory implies that the average response should fall between the stimulus value and that of the average of the potential responses.

The present results provide conflicting evidence with respect to the more applicable theory. On the one hand, the average response under the Unfavorable condition was clearly consistent with adaptation-level theory since it was reliably different from the value of the relevant Initial List in the direction implied by the theory. On the other hand, the average response under the Favorable condition was not reliably different from the value of the relevant Initial List, and this outcome is consistent with the generalization theory. However, the obtained difference, though not significant, was in the direction implied by the adaptation-level theory. Taken together, these two essentially independent results indicate that both theories require consideration when further studies are planned in this area.

Comparison with Psychophysical Methods

The similarity of the present results to those involving judgment of physical qualities, e.g., weight, loudness, is impressive because of the radical difference in stimulus quality, and because the present procedures were substantially different from those observed in conventional psychophysical experiments. For example, although the traits were scaled for favorableness, and the judgments were demonstrated to have been determined by this factor, subjects did not deliberately judge on this basis. In fact, most subjects were found to be unaware of it. This lack of awareness of the basis of judgment is in contrast to conventional psychophysical situations in which the subject judges stimuli with respect to an explicitly agreed upon dimension. Also, although the Check List was treated as a series of comparison stimuli, the subjects were not actually required to compare the items one by one with their initial impressions as in the usual psychophysical experiment. Nevertheless, judgments obtained under these relatively loosely controlled conditions were analyzable by means of Helson's technique, and the results were interpretable in terms of the adaptation-level theory.

Impressions of Persons

Favorableness of personal traits was shown by the present results to be an important determinant of behavior in the impressions of persons situation. Therefore, the effects of the words WARM and COLD in previous studies can be explained in terms of the substantial difference in favorableness (warm, 2.2; cold, 5.1) between the two terms, rather than by the "central trait" theory advanced by Asch (1946). More specifically, a difference in impressions is produced by a difference in the favorableness of the initial trait-names. Following this same reasoning one can explain Asch's (1946) finding that, in contrast to WARM and COLD, POLITE and BLUNT had little differential effect on traits chosen on the basis of an initial impression. Assuming, as before, that the favorableness dimension was dominant, we hypothesized that the difference between POLITE and BLUNT with respect to the favorableness dimension is smaller than that be-

tween WARM and COLD. In order to test this hypothesis, 52 subjects from the same population as that sampled in the main experiment rated 18 trait-names including WARM, COLD, POLITE, and BLUNT, on a seven-point scale of favorableness.

The mean difference between the ratings of COLD and WARM proved to be 4.0 scale intervals, while the corresponding difference between the ratings of POLITE and BLUNT was only 2.5 intervals. The difference of 1.5 scale intervals between these two values was significantly different from zero beyond the 1% level, and it indicates that POLITE and BLUNT are less different from each other with respect to favorableness than are WARM and COLD. Therefore, it is reasonable to expect that there will be less difference between traits chosen under "polite" as compared to "blunt" conditions than under "warm" as compared to "cold" conditions.

SUMMARY

Twenty-four subjects were given an Initial List of three very favorable trait-names, while another 24 subjects were given a different Initial List consisting of three very unfavorable trait-names. Each subject was instructed to form an impression of a fictitious person, and then to select from a check list four other trait-names appropriate to the initial impression. The Check List consisted of 25 trait-names that had been scaled previously on the favorableness dimension. Trait-names selected under the Favorable condition were consistently and considerably more favorable than those selected under the Unfavorable condition. In addition, under the Unfavorable condition the average value of selected trait-names was consistently more favorable than the average value of the associated Unfavorable Initial List. Under the Favorable condition, however, an analogous deviation toward unfavorableness was not statistically reliable.

The relationship between verbal and quantitative connotation was discussed in relation to Osgood's dimension theory of connotation, mediated generalization, and Helson's adaptation-level theory. In addition, Asch's impressions of persons situation was reinterpreted in terms of the dimension theory of connotation. Finally, it was shown that the lesser effect of POLITE versus BLUNT as compared to that of WARM versus COLD is explainable in terms of quantitative connotation.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

HELSON, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HELSON, H., DWORKIN, R. S., & MICHELS, W. C. Quantitative denotation of common terms as a function of background. *Amer. J. Psychol.*, 1956, **69**, 194–208.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KELLEY, H. H. The warm-cold variable in first impressions of persons. *J. Pers.*, 1950, **18**, 431–439.

MOSIER, C. I. A psychometric study of meaning. *J. soc. Psychol.*, 1941, **13**, 123–140.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. Illinois Press, 1957.

RAZRAN, G. H. S. Salivating and thinking in different languages. *J. Psychol.*, 1935, **1**, 145–151.

RIESS, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.

SEARS, F. W. *Principles of physics*. Vol. 1. Cambridge, Mass.: Addison-Wesley, 1946.

Webster's new world dictionary of the American language. New York: World, 1951.

WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.

(Received May 24, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 598-604

PSYCHOTHERAPEUTIC AND OTHER HOSPITAL TREATMENT CRITERIA: THE DILEMMA

RALPH P. FORSYTH[1] AND GEORGE W. FAIRWEATHER[2]
Veterans Administration Hospital, Palo Alto, California

In recent years, psychotherapists and researchers have used many different assessment devices as criteria for the adjustment of the mentally ill patient. Because of this diversity of criteria, most studies evaluating different neuropsychiatric treatments are not comparable even when similar patient populations have been used. The most common indices of adjustment are hospital behavior rating scales, personality inventories, projective devices, and follow-up ratings. Two recent studies (Fairweather, Simon, Gebhard, Weingarten, Holland, Sanders, Stone, & Reahl, 1960; Lasky, Hover, Smith, Bostian, Duffendack, & Nord, 1959), in which a small number of diverse measures were used, have indicated that these measures are not related to one another, a finding which cautions against generalizing from test responses or other hospital behaviors to community adjustment. This study investigates the relationships among historical indices, personality inventory test responses, projective devices, within-hospital behaviors, and follow-up ratings in order to arrive at an empirical criterion for assessing patient adjustment.

METHOD

The data used in this study were collected at a large Veterans Administration hospital in order to compare the relative efficiency of four psychotherapeutic programs for three different patient groups. The results of these comparisons were given in a monograph (Fairweather et al., 1960) which presents a detailed account of the design, methods of data collection, and development of instruments.

In the aforementioned study, each of the 96 male veteran patients was tested on a variety of psychological instruments during his last week of hospitalization. One-third of the patients were nonpsychotics—behavior disorders and neurotics; one-third were long-term (chronic) psychotics—mostly schizophrenics who had been hospitalized for more than one year; the remaining third were short-term (acute) psychotics—mostly schizophrenics who had been hospitalized less than one year.

The tests administered were the Thematic Apperception Test (TAT), *Q* sorts of both ideal and real self, the Minnesota Multiphasic Personality Inventory (MMPI), and the Holland Vocational Personality Inventory (HVPI) (Holland, 1958). Also, ratings on both job and ward behavior were obtained. For those 48 patients participating in group therapy, a group therapy rating was completed every 30 days (Finney, 1954). A record was kept of time in treatment and tranquilizing drugs received while in the hospital. In addition, historical information was collected for each patient in the study, and 23 items were selected as possible prognostic indices. Finally, 6 months after each patient's discharge, a questionnaire was completed by an individual with whom the patient was living at that time.

Thus, the data for this study are composed of 66 items or scales, each of which was dichotomized to reflect the adaptive-maladaptive dimensions of the variable. The purpose is to determine the relationships among these 66 frequently used clinical variables in order to arrive at a meaningful set of criteria for evaluating patient adjustment.

RESULTS

A 66 × 66 tetrachoric correlational matrix was computed and a cumulative communality (CC) cluster analysis performed (Tryon, 1958). Factoring by the key cluster method was accomplished on an electronic computer. Seven orthogonal dimensions were found.[3] Seven oblique clusters were located on these dimensions. In order to discover the relationship between each variable and the seven cluster domains, its unaugmented oblique loadings with the domains were computed (Tryon, 1958). Variables with unaugmented oblique loadings of .50 or over with the clusters are presented in Table 1.

Cluster 1 is the MMPI domain of self-evaluation; it includes all the scored MMPI scales (except *L*, *K*, *Pd*, *Mf*, *Ma*, and Welsh *R*),

[1] Now at the University of North Carolina.

[2] The authors are deeply indebted to Robert C. Tryon for his invaluable advice and interest in all phases of this research. Also, we are grateful to Ralph Simon, Mildred E. Gebhard, Eric Weingarten, John L. Holland, Richard Sanders, Gideon B. Stone, and Joseph E. Reahl for their aid in collecting the data for this study.

[3] Factoring was terminated in accordance with the objective criteria set forth by Tryon (1958).

TABLE 1
THE SEVEN CLUSTERS

Cluster 1	
a. Adaptive MMPI-CS[b]	(.98)[a]
b. Low MMPI-*F*	(.92)
c. Low MMPI-*Pt*	(.89)
d. Low MMPI-*SIM*[c]	(.88)
e. Low MMPI-*Pa*	(.86)
f. Low MMPI-*Sc*	(.82)
g. Low MMPI-Welsh *A*[d]	(.82)
h. High ideal-self *Q* sort correlation	(.82)
i. Low MMPI-*D*	(.80)
j. High MMPI-*Es*	(.75)
k. Low MMPI-*Hy*	(.73)
l. Low MMPI-*Hs*	(.64)
m. Few previous arrests	(.51)
Cluster 2	
a. High HVPI 5 (Verbal Activity)	(.94)
b. High HVPI 12 (Responds like others)	(.91)
c. High HVPI 4 (Conformity)	(.87)
d. High HVPI 6 (Emotional)	(.87)
e. High HVPI 8 (Aggressiveness)	(.87)
f. High HVPI 3 (Responsibility)	(.86)
g. High HVPI 9 (Mf-masculinity)	(.73)
h. High HVPI 2 (Intellectuality)	(.72)
i. High HVPI 1 (Physical Activity)	(.63)
j. High HVPI 10 (Status)	(.53)
Cluster 3	
a. Older age at first hospitalization	(.96)
b. Older onset age	(.96)
c. No service connected pension	(.69)
d. Nonpsychotic diagnosis (upon admission)	(.60)
e. Few total weeks of previous hospitalization	(.57)
f. Older age upon admission	(.52)
g. Fewer number of previous hospitalizations	(.51)
Cluster 4	
a. Maladaptive follow-up drinking	(.86)
b. High HVPI 9 (Mf-masculinity)	(.77)
c. Nonpsychotic diagnosis upon admission	(.64)
d. Low HVPI 7 (Control)	(.60)
e. Few total weeks of previous hospitalization	(.57)
f. Adaptive Finney scale score	(.55)
g. Adaptive ward behavior	(.55)
h. Low HVPI 11 (Infrequency)	(.53)
i. Low MMPI-Welsh *R*[d]	(.51)
Cluster 5	
a. High service discharge rank	(.83)
b. High gain in service rank	(.75)
c. Low HVPI 11 (Infrequency)	(.63)
d. More education completed	(.59)
e. Few service courts-martial	(.56)
f. High socioeconomic status	(.56)
Cluster 6	
a. Follow-up, employed full-time	(.96)
b. Follow-up, well most of time	(.73)
c. Follow-up, was not rehospitalized	(.73)
d. Adaptive Finney scale score	(.66)
Cluster 7	
a. Few previous arrests	(.82)
b. Few jobs held in past	(.64)
c. Few service courts-martial	(.61)
d. Maladaptive job behavior in hospital	(.60)

[a] Unaugmented cluster loadings in parentheses.

[b] In addition to the usual MMPI scales, configural interpretations of the MMPI and HVPI were obtained. George S. Welsh and John L. Holland, experts on each of these tests, were asked to place each test profile into an adaptive or maladaptive group in terms of the expected follow-up adjustment of each patient. These two scales were designated, respectively, as an MMPI Configural Scoring (MMPI-CS) and, similarly, HVPI-CS.

[c] *SIM* is the MMPI scale of Inner Maladjustment (Simon, 1952).

[d] MMPI Welsh *A* and *R* refer to Welsh's (1956) Anxiety and Repression scales.

the *Q* sort correlation between ideal and real self,[4] and number of arrests prior to hospitalization. All the measures in Cluster 1 reflect the adaptive ends of their scales; i.e., if a patient scores adaptively on one MMPI scale he tends to score adaptively on all of them. Conversely, of course, the same is true for maladaptive scores.

Cluster 2 is a domain of vocational preferences, the HVPI. All the HVPI scales except HVPI 7 (Control), 11 (Infrequency), and the HVPI-CS have high loadings. As in Cluster 1, subjects who respond adaptively on one HVPI scale tend to score adaptively on all HVPI scales. Holland (1958) discusses the personality correlates of this test.

Cluster 3, containing seven historical items, refers to the neurotic-psychotic dimension. It seems to differentiate those patients who are currently hospitalized with neurotic or behavior disorders from those hospitalized with psychotic reactions. These older, nonpsychotic veterans had fewer previous hospitalizations, fewer total weeks of hospitalization in the past and, of course, were older at time of onset of illness and when first hospitalized. They also tended not to receive a service connected pension since their breakdown usually did not occur in service. This appears to be a nonpsychotic vs. chronic-psychotic cluster.

[4] These correlations were computed for each patient and then dichotomized at the median. Those correlations higher than average were considered adaptive, those lower, maladaptive.

Cluster 4 seems to be a psychopathic index. These patients, also generally nonpsychotic, adjust well on the hospital ward, interact well in group psychotherapy (Finney scale), and have rarely been hospitalized in the past. They incline to alcoholism when leaving the hospital. The psychological test scales in this cluster indicate that patients high on this cluster chose more masculine occupations (HVPI 9–Mf), were "impulsive and undercontrolled" (HVPI 7), chose the more usual occupations (HVPI 11–Infrequency), and were low in repression" (MMPI Welsh *R*).

Cluster 5 seems to be an achievement domain. These patients are better educated and concomitantly have higher occupational status as well as service discharge ranks. They also have advanced more service ranks and have had few courts-martial while in the service. Their lower score on HVPI 11 (Infrequency) indicates that these patients selected the more usual occupational choices on this inventory.

Cluster 6 refers to the domain of posthospital adjustment. It contained three follow-up items (employment, rehospitalization, and well most of the time) and the Finney (1954) group psychotherapy scale. All four cluster together in the adaptive directions, i.e., is employed, out of hospital, well most of the time, and scores high on the therapy scale.

Cluster 7 seems to sample hospital job maladjustment. High scores describe individuals who have historically conformed to societal demands but have difficulty in adjusting to a routine hospital job. This type of patient has had few arrests, few jobs, and few service courts-martial, but he seems unable to function adequately in his unchallenging hospital job.

Estimates of the theoretical intercorrelations among the seven cluster domains, as well as the domain validities and communalities, are shown in Table 2. The correlations between actual cluster scores (the theoretical intercorrelations "knocked down" by their invalidities, Tryon, 1958, p. 29) are also reported in parentheses in Table 2. The domain validities are so high that there is little difference between the intercorrelations of the cluster domains and cluster scores.[5] Table 2 shows the relative independence of the clusters; only three cluster domain intercorrelations reach .30. Both the adaptive HVPI cluster (2) and the "nonpsychotic" cluster (3) correlate in this magnitude with the "psychopathic" cluster (4). The adaptive MMPI cluster (1) and the historically adaptive cluster (7) are similarly related. Three other cluster intercorrelations reached a magnitude of .26 or .27. Both the psychopathic (4) and historically adaptive (7) dimensions are negatively related to the achievement factor (5). Also, the adaptive MMPI cluster (1) and the adaptive follow-up cluster (6) are positively related to a moderate degree.

Twenty of the 66 variables were not included in any of the seven clusters. Four were test variables (MMPI *L* and *Mf*, HVPI CS, and the TAT[6]). Seven were historical items (race, received tranquilizing drugs in previous hospitalizations, reared by both parents, percentage of time gainfully employed in last 10 years, weekly salary of last job, type of discharge, marital status). Two hospital variables (time in treatment, and received tranquilizing drugs during current hospitalization) and two follow-up items (has some, or no close friends, and frequently or rarely talks to people with whom he lives) were also not included in any cluster.

Seven of the scales were in two of the seven clusters. These were: number of weeks of previous hospitalization (Clusters 3 and 4); number of arrests (1 and 7); number of service courts-martial (5 and 7); type of admission diagnosis (3 and 4); group psychotherapy scale (4 and 6); HVPI 9–Mf (2 and 4); and HVPI 11–Infrequency (4 and 5).

Table 3 presents the original tetrachoric correlations between the five within-hospital treatment variables and the six follow-up items. High scores for each variable are adaptive so that a positive correlation indicates

[5] The values in parentheses are approximations of the product-moment correlations between cluster series because the tetrachoric correlations used in the analyses are themselves approximations of the product-moment correlations between the individual tests. The approximations should be fairly accurate, however, since the dichotomizing of most variables was of an observed continuous distribution.

[6] Eight selected TAT cards (1, 2, 4, 9, 13MF, 7BM, 6BM, and 12M) were given to each patient and then scored, without identifying information, for adaptivity in expression of affect, level of plot development, and degree of expressed interpersonal relationships.

TABLE 2

INTERCORRELATIONS BETWEEN THE CLUSTER DOMAINS (Oblique Factors)

	MMPI	HVPI	Neurotic-Psychotic	Psychopathic	Achievement	Follow-up	Conformity
MMPI	1.00 (.98)[a]	−.03 (−.03)	.12 (.10)	.03 (.03)	−.07 (−.06)	.26 (.25)	.36 (.33)
HVPI	−.03 (−.03)	1.00 (.97)	.07 (.06)	.32 (.30)	.00 (.00)	.12 (.12)	.18 (.17)
Neurotic-Psychotic	.12 (.10)	.07 (.06)	1.00 (.90)	.31 (.29)	−.20 (−.18)	−.04 (−.03)	.09 (.07)
Psychopathic	.03 (.03)	.32 (.30)	.31 (.29)	1.00 (.93)	−.27 (−.25)	.07 (.07)	.04 (.03)
Achievement	−.07 (−.06)	.00 (.00)	−.20 (−.18)	−.27 (−.25)	1.00 (.88)	.17 (.15)	−.27 (−.24)
Follow-up	.26 (.25)	.12 (.12)	−.04 (−.03)	.07 (.07)	.17 (.15)	1.00 (.94)	.07 (.06)
Conformity	.36 (.33)	.18 (.17)	.09 (.07)	.04 (.03)	−.27 (−.24)	.07 (.06)	1.00 (.89)

DOMAIN VALIDITIES AND COMMUNALITIES
(Predictable Variance)

	Clusters						
	1	2	3	4	5	6	7
Communality of cluster scores	.98	.97	.90	.93	.88	.94	.89
Domain validity of cluster score	.99	.98	.95	.96	.94	.97	.94

[a] Intercorrelations between cluster scores in parentheses.

TABLE 3

TETRACHORIC CORRELATION MATRIX OF HOSPITAL VARIABLES AND FOLLOW-UP ITEMS

	Ward Behavior	Job Behavior	Drugs (on)	Time in Treatment	Finney Scale	Not Rehospitalized	Employed Full Time	Well most of Time	Talks to Others	Has Friends	Not a Problem Drinker
Within hospital assessments											
Ward behavior		.12	.04	−.26	.50	.04	.28	−.32	.23	.14	−.47
Job behavior	.12		−.10	.22	−.14	.02	−.06	.08	−.22	.11	−.41
Drugs (on)	.04	−.10		.10	.22	−.02	−.07	−.11	−.19	−.19	.18
Time in treatment (long)	−.26	.22	.10		−.20	−.30	.08	−.04	−.23	.07	.24
Finney scale	.50	−.14	.22	−.20		.42	.75	.09	.20	.29	−.54
Follow-up											
Not rehospitalized	.04	.02	−.02	−.30	.42		.66	.59	.09	.12	.29
Employed full time	.28	−.06	−.07	.08	.75	.66		.57	.30	.46	.14
Well most of the time	−.32	.08	−.11	−.04	.09	.59	.57		.46	.50	.60
Talks to others	.23	−.22	−.19	−.23	.20	.09	.30	.46		.51	.06
Has friends	.14	.11	−.19	.07	.29	.12	.46	.50	.51		−.08
Not a problem drinker	−.47	−.41	.18	.24	−.54	.29	.14	.60	.06	−.08	

Note.—Enclosed area contains Cluster 6.

agreement between theoretical levels of adaptation. Table 3 points out the general lack of agreement between the within-hospital and follow-up measures. Only the Finney (1954) scale, of the within-hospital measures, was related with any magnitude to more than one follow-up measure.

The follow-up items tended to be relatively more independent than might be expected. Follow-up items concerning talking to other people and problem drinking tended to be only slightly positively related to the others. And, curiously enough, inspecting the original correlational matrix one finds that the follow-up drinking item was related much more frequently to historical and psychological test

scales than any other follow-up item. These were: .49 with few previous weeks hospitalized, .49 with past drinking problem, .62 with nonpsychotic diagnosis, −.51 with MMPI *L*, −.44 with MMPI *K*, .67 with MMPI *Ma*, and −.58 with HVPI 9 (Mf).

DISCUSSION

This cluster analysis of 66 indices of patient adjustment emphasizes the relative independence of frequently used criterion measures. For the hospitalized patient population used in this study, the common criteria of ward and job behavior ratings, usually given highest weight in hospital discharge policy, are relatively unrelated to any follow-up area of adjustment. Further, prognostic indices such as past achievement, projective devices, and personality inventories are generally not related to community adjustment. Follow-up items themselves show a wide range of correlations, i.e., follow-up full time employment correlates .66 with remaining out of the hospital, but only .12 with number of friends.

In view of the heterogeneity of criteria, it appears that studies attempting to evaluate hospital treatments can only be generalized when a common criterion is chosen. Inasmuch as hospitalization for mental illness, whether voluntary or committed, represents a failure of the self-maintenance system, it seems only logical that an evaluation of hospital treatments should be made in the areas of community living representing this failure. An approximation to such a criterion is found in the follow-up cluster which involves simple adjustmental indices of rehospitalization, work, and an estimate of the former patient's current degree of illness.

While the follow-up cluster would be valuable in evaluating hospital treatments, it would seem to have little merit for assessing outpatient treatments. Hollingshead and Redlich (1958) found that most analysands are members of higher socioeconomic groups and hence would be out of the hospital, almost universally employed, and characterized by a limited range pathology. Further, outpatients frequently enter treatment voluntarily because of subjective discomfort rather than because of any readily observable breakdown of the self-maintenance system. Thus, it appears that a social criterion for the evaluation of outpatient treatments deals with such a small range of the continuum that comparisons would be logically meaningless. Moreover, by the very nature of many outpatient populations, subjective evaluations are essential. It seems clear that a universal criterion for evaluating treatment success is nonexistent and situational criteria, at least within or outside hospitals, are necessary.

The only hospital measure found in the follow-up cluster is the Finney (1954) Group Therapy Rating. The high loading of this scale with follow-up items may be accounted for by two independent but equally important variables. First, the scale itself comprises items related to readily observable behaviors in emotionally loaded interpersonal situations. This is in direct contrast to ward and job behavior ratings, which usually involve conformity behavior with less emotional involvement. Second, the therapy rating scale was completed every 30 days, and the rating used in this analysis was the one prior to discharge.

Figure 1 compares all patients who received group therapy when divided into the most and least adaptive follow-up groups. This figure demonstrates that follow-up adjustment is most accurately predicted by the last of a series of observations in the therapeutic situation. It also suggests that there may be a point in time when therapy should be terminated, if progress in interpersonal behavior has not occurred. In Figure 1, the most adaptive follow-up group becomes significantly differentiated from the least adaptive at 120 days (t of 3.26 with 25 df, $p < .01$).

It is not surprising, then, that ratings made once or twice during the patients' hospitalization are not highly related to follow-up. Further, the therapy rating loading on the follow-up cluster suggests that social-psychological indices of behavior in close interpersonal contexts within the hospital is the area of measurement most deserving of further research and expansion.

Future studies must also take cognizance of the significantly poorer follow-up adjustment of the long-term psychotics when compared with short-term psychotics and nonpsychotic groups. A previous study (Fairweather et al., 1960) clearly showed that schizophrenics cannot be compared without the amount of patients' previous hospitalization being con-

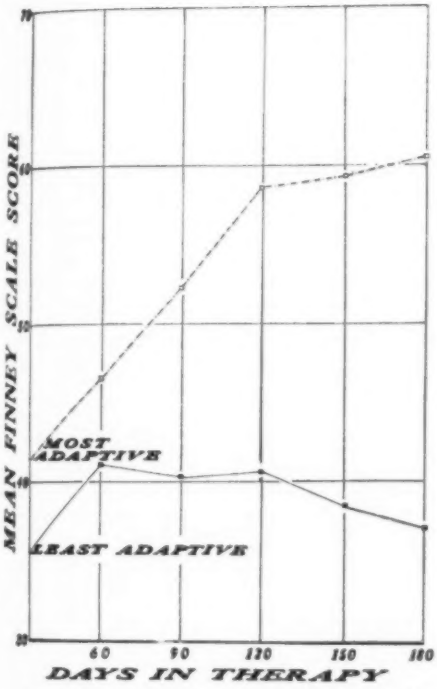


FIG. 1. Group therapy ratings for patients who demonstrated most and least adaptive community adjustment. (All nine follow-up items were dichotomized into adaptive and maladaptive alternatives, with adaptive alternatives receiving a score of one and maladaptive zero. The community adjustment score for a patient was the sum of the nine items. The distribution of community adjustment scores was dichotomized at the median and those above the median were labeled most adaptive and below the median least adaptive.)

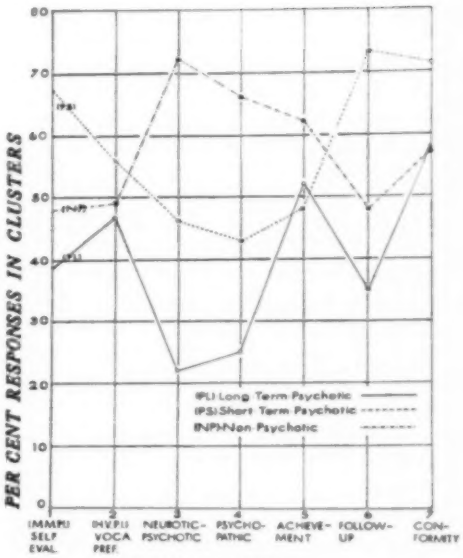


FIG. 2. Comparison of the three diagnostic groups on the seven clusters. (The ordinate, Per Cent Responses in Clusters, is the ratio of hits in a cluster to total number of possible hits. Because each variable is dichotomized, only .5 of any distribution can be in a cluster. Thus, on the MMPI cluster with 13 scales and 32 patients in a diagnostic group, there are 13 × 32 or 416 possible hits. Figure 2 shows that the Ps group hit 283 of the possible 416 or 68%, while the Pl group hit 162 of the 416 or 39%.)

trolled because short-term psychotics demonstrate significantly more adaptive follow-up than long-term psychotics, irrespective of treatment. It may be that norms for each group will have to be independently established.

The three diagnostic groups (nonspychotics, short-term, and long-term psychotics) may also be compared on the seven clusters found in this study. Figure 2 shows that the acute psychotics scored much more adaptively than the other groups on the MMPI cluster (1), the follow-up cluster (6), and the historically adaptive cluster (7), while looking slightly better on the HVPI cluster (2). The nonpsychotic group, of course, has high scores on the nonpsychotic cluster (3), the psychopathic (4), and achievement (5) clusters. The chronic psychotics were least adaptive on the test clusters (1 and 2) and the follow-up cluster (6); they were also very low on Clusters 3 and 4, which is indicative of their chronicity and lack of psychopathy.

As previously mentioned, the two test clusters are relatively orthogonal to the follow-up cluster. The number of highly interrelated scales in Clusters 1 and 2 suggest that the major portion of the variance of both tests may be explained by one general variable which, in this case, is unrelated to the to-be-predicted follow-up response. In this regard, the MMPI configural scoring was the best single measure of the MMPI factor, but had little or no relationship to follow-up. Thus, the MMPI configural scoring predicted the rest of the MMPI scales well but had little association with hospital behavior or follow-up. This finding questions the value of profile analysis over single scale scores. The very high loading

of the *Q* sort correlation in Cluster 1 suggests that self-evaluation items are common to these two forms of measurement.

The fact that much-quoted and frequently-used tests such as the TAT failed to appear in any cluster demonstrates the problem of relating fantasy and self-evaluation reports to each other and to complex behaviors. It seems that much higher correlations need to be demonstrated before psychologists use such tests as decision making instruments where complex behaviors are the predicted response.

The lack of relationship between length of time in treatment and follow-up corroborates a recent study by Lasky et al. (1959). Similarly the use of tranquilizing drugs seems to have little relation to posthospital adjustment.

SUMMARY

Sixty-six frequently used clinical variables, including historical data, personality inventories, projective devices, hospital behaviors, and follow-up ratings, were collected from 96 male NP patients from a large Veterans Administration hospital. These variables were intercorrelated and cluster analyzed. Seven clusters accounted for all but negligible variance. One of these clusters, which included three follow-up items, was proposed as a relatively meaningful criterion which could be used to evaluate psychotherapeutic and other hospital treatments for this type of hospitalized population. The lack of relationship between psychological tests and follow-up was discussed. The importance of measuring interpersonal interactions within the hospital was suggested as a meaningful avenue of research because of the promising prognostic results obtained with the Finney (1954) group psychotherapy scale.

REFERENCES

FAIRWEATHER, G. W., SIMON, R., GEBHARD, M. E., WEINGARTEN, E., HOLLAND, J. L., SANDERS, R., STONE, G. B., & REAHL, J. E. Relative effectiveness of psychotherapeutic programs: A multicriteria comparison of four programs for three different patient groups. *Psychol. Monogr.*, 1960, **74** (5, Whole No. 492).

FINNEY, B. C. A scale to measure inter-personal relationships in group psychotherapy. *Group Psychother.*, 1954, **7**, 52–66.

HOLLAND, J. L. *Manual for the HVPI.* Palo Alto: Consulting Psychologists Press, 1958.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness: A community study.* New York: Wiley, 1958.

LASKY, J. J., HOVER, G. L., SMITH, P. A., BOSTIAN, D. W., DUFFENDACK, S. C., & NORD, C. L. Posthospital adjustment as predicted by psychiatric patients and by their staff. *J. consult. Psychol.*, 1959, **23**, 213–218.

SIMON, R. An investigation of the relationship between several measures of self-insight and adjustment, intelligence and personality characteristics. Unpublished doctoral dissertation, Syracuse University, 1952.

TRYON, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt*, 1958, **18**, 3–35.

WELSH, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine. Measmt.*, Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received May 25, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 605-610

INTRAJUDGE CONSISTENCY IN RATINGS OF HETEROGENEOUS PERSONS[1]

CECILY F. GROSS
Teachers College, Columbia University

In an analysis of components of accuracy scores concerning social perception, Cronbach (1955) noted that observer bias merits attention beyond its role as a source of constant error. He suggested that the mean and standard deviation of the ratings by a judge (*J*) of a number of other persons (*O*s) constitute an implicit personality theory. Although research on specific determinants of impression formation has demonstrated the relevance of a number of attitudes and cognitive modes of the judge, attributes of the stimulus, and characteristics of the social interaction between them, evidence for a reliable response set with generality over heterogeneous others has not been clear-cut.

Unmeasured specific biases of *J* toward *O* seem likely to have increased the consistency of judgment reported in some studies. In the research of Crow and Hammond (1957), for example, where the *O*s were hospital patients and the *J*s medical students, a factor may have been the *J*s' attitudes related to their role vis-à-vis patients. Likewise, in Alfert's (1958) study, consistency may have been inflated by the fact that, although unknown to one another at the time, the *J*s were nonetheless rating their own classmates.

More satisfactory evidence of individual consistency of response toward heterogeneous other persons has been shown in studies by Beier, Smock, Izard, and Tougas (1957) and Izard (1959) in which stimulus photographs were selected from a pool of over 1,000 taken of males and females of widely differing ages. Yet a set of just as carefully selected photos led to inconclusive results in a study by Campbell and Burwen (1956). Although these investigators found highly significant between judge differences within a group of WAFs, they found no significant respondent effects within any of their three groups of male Air Force personnel. It also seems noteworthy that, in contrast to their apparent difficulty in demonstrating consistency of response to heterogeneous others on the part of three of the four groups studied, Campbell and Burwen obtained impressively high reliability coefficients for the ratings of specific age and sex subgroups of *O*s, despite proportionately small *N*s. Perplexingly, the male subjects in a study by DeSoto, Kuethe, and Wunderlich (1960), responding to the same stimulus pictures used by Campbell and Burwen, showed marked intrajudge consistency.

The second parameter of Cronbach's (1955) "implicit personality theory," the standard deviation of a judge's responses to a large number of heterogeneous others, has been the topic of very little research other than the previously mentioned study by Crow and Hammond (1957) in which reliable intrajudge variability was found. The question whether judgmental consistency is general across descriptive dimensions has also received little attention.

The purpose of the present study was to investigate the existence and relative power of a unitary attitude toward heterogeneous others within and across descriptive dimensions, employing Cronbach's (1955) operational definition. The major hypothesis was that on any given descriptive dimension, individual judges tend to be consistent in their ratings of a group of stimulus persons. Since individual differences in intrajudge consistency can be assumed to obtain across all traits being judged, the second hypothesis was that the tendency to consistency on one trait is positively related to consistency on all. Subordinate hypotheses were that no differences would be

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Teachers College, Columbia University, 1960. The writer wishes to express gratitude to the members of the Dissertation Committee, Robert L. Thorndike and Edward J. Shoben, Jr., and particularly to thank the Chairman, Joel R. Davitz, for his stimulating encouragement. Deep appreciation is also extended to Rosedith Sitgreaves for her generous help.

The graciousness of H. Schmierer of Pratt Institute is gratefully acknowledged.

found related to the sex of the judge nor the sex of the stimulus person.

METHOD

Judges

The subjects serving as judges were 35 female and 27 male freshmen and sophomores enrolled in a general psychology course at Pratt Institute. The average age of the women was 18.7 and that of the men 22.4 years. The statistically significant 4-year difference is accounted for by the previous military experience of some of the men. Of the women, 24 were majors in art teacher education, the remainder being majors in fashion, interior design, and art. Of the men, 19 were architecture students, with the rest distributed among art teacher education, industrial engineering, mechanical engineering, and fashion design.

Procedure

The design required the 62 *J*s to view and rate a number of *O*s on various scales. The *J*s were shown a series of 30 30-second motion pictures in each of which a single stimulus person was seen in a pattern of standardized, relatively neutral activity. The sequences, taken outdoors, showed a full view of the *O* walking a distance of approximately 10 feet to a bench and then sitting down. During the last 10 seconds of film, the camera panned in for a close-up.

The 30 stimulus persons (*O*s) ranged in age from the twenties to the forties, five persons of each sex falling in each of the three decades. Because the purpose of the experiment was to investigate the generality of anticipatory set, heterogeneity among the superficial characteristics of the *O*s was essential. However, the selection of *O*s was restricted to Caucasians, and their dress to street clothes in order to avoid contamination by racial or socioeconomic attitudes. The order in which the *O*s appeared on the screen was randomly determined, except that no more than two *O*s of the same sex were shown in succession.

The rating scales upon which the *J*s' impressions were recorded were derived from the Guilford-Zimmerman Temperament Survey. The seven factors on which the scales were based were general activity, rhathymia, ascendance, sociability, emotional stability, masculinity-femininity, and friendliness. These traits were chosen for their importance and relative independence as dimensions of personality according to Guilford's (1949) research. Furthermore, since they occur commonly in personality description, their use was presumed to have the advantage of appearing plausible to the *J*s. A brief group of synonyms and behavioral statements was used to define, illustrate, and anchor each end of the seven bipolar scales.

In order to prevent the *J*s from seeing their previous ratings, the rating scales were presented in booklet form, with a set of seven scales for the rating of one *O* to each page. An attempt was made to eliminate order and context effects by varying the order of presentation and the left to right position of each scale from page to page.

The data were collected at a single meeting of the class. The following instructions, reproduced on the cover of the booklets, were read aloud:

> Whenever we see a stranger for the first time, we get an idea about what kind of person he is. In a few minutes we are going to show you movies of a number of men and women, one at a time. We are interested in finding out how they impress you. We'd like you to tell us your hunches about what each person is really like by rating him on certain traits.
>
> Each one of the pages in this booklet is to be used to describe one of the persons you will see in the movie. On each page you will find seven pairs of traits, along with a matching set of seven rating scales. The scales run from the extreme of each trait all the way to its opposite. After you see each person on the screen, rate him on ALL SEVEN of the rating scales on the page. Do this for each scale by putting an "x" in the middle of the *one* box that fits your impression best.
>
> You will have only 30 seconds to look at each person, and then the projector will be stopped for you to make your ratings. So don't take time to puzzle over your ratings or worry if you are not too sure. Your snap judgment is exactly what we want!

Supplementary informal verbal instructions were designed to encourage frankness and to reduce the *J*s' tendency to give the socially acceptable response. So that tendencies to evasiveness might also be minimized, the *J*s were asked to try to avoid using the midway point of the scales unless it seemed precisely to fit their impression of the *O*. After a practice trial to be certain that the *J*s understood the use of the scales, the films were run, the lights being turned on between each section of film to permit the *J*s to record their ratings of the *O* just viewed.

The statistic used as a coefficient of consistency was the intraclass correlation (r') (Haggard, 1958) obtained by analysis of variance of repeated measures on subjects who thus serve as their own controls. The significance level of this coefficient is determined by the *F* test. Although the significance of the difference between two r''s can be estimated only by examining the degree of overlap between the confidence intervals of a pair of r''s, the frequency with which differences fall in a predicted direction is amenable to the nonparametric sign test (Siegel, 1956).

RESULTS

Intradimensional Consistency

Testing Hypothesis I, that individual judges tend to consistency of response toward all *O*s within each rating dimension, involved separate analyses of variance of the *J*s' ratings on each scale. Statistically significant between judge differences at better than the .01 level appeared for all seven of the scales. The Between Judge Mean Square (BJMS) for each scale was then used to obtain an intraclass correlation as the summary index of consistency of the

TABLE 1

COEFFICIENTS OF CONSISTENCY FOR 62 *Js* JUDGING 30 *Os* ON EACH OF SEVEN SCALES (σ_o^2 = WMS)

Scale	Single Item	Overall
I	.049	.61
II	.057	.64
III	.011	.24
IV	.032	.49
V	.082	.73
VI	.092	.75
VII	.036	.53
All Scales	.162	.57

TABLE 2

COMPONENTS OF VARIANCE FOR 62 *Js*' JUDGMENTS ON EACH OF SEVEN SCALES

Scale	Judges		Objects		Error		Total
	Variance	%	Variance	%	Variance	%	
I	.15	5	.55	21	1.91	74	2.6
II	.16	6	.23	8	2.30	86	2.7
III	.05	1	.71	24	2.20	75	2.9
IV	.11	4	.78	28	1.90	68	2.8
V	.22	8	.39	15	1.93	77	2.5
VI	.24	10	.28	11	1.98	79	2.5
VII	.09	4	.45	21	1.59	75	2.1

Js' responses on that particular trait. Since this study was concerned essentially with intrajudge consistency as compared with all other sources of variation, known or unknown, the appropriate error term in computing r' was the Within Mean Square (WMS) obtained after the Between Judge Sum of Squares was subtracted from the Total Sum of Squares. In Table 1 are the reliability coefficients of the seven scales, both for ratings of a single *O* and for the total of 30 ratings.

It can be seen that if the series of stimulus persons is regarded as a test, the reliability of a single item (*O*) on any one scale is practically nil; while the reliabilities for whole "tests" of 30 items per scale range from .24 to .75.

Because the between objects differences in these seven analyses of variance were highly significant, r'''s were also computed in which the the error terms were the Residual Mean Squares (RMS) remaining after Sums of Squares for Objects and Judges both were removed. For all scales, the RMS coefficients were larger than those based on the WMS. The increment in the intraclass correlations following the removal of object effects can be attributed to the fact that the BOMS ranged from 2 to 12 times as large as the BJMS. The relative slightness of the increments, despite the magnitude of these ratios, probably can be accounted for by the large proportion of error variance appearing in the analysis of the components of variance for each scale shown in Table 2. So great a proportion of error variance is only to be expected in a design providing insufficient information for veridical judgment. However, the finding that the stimuli, despite their seeming scantiness, contributed from 8% to 28% of the total variance seems surprising indeed in the context of expectancy theory. In sharp contrast, the proportion of variance attributable to observer bias is miniscule, although the research method had been selected because of the likelihood that it would maximize judge factors.

In summary, the data show the presence of intrajudge consistency of response within each scale. However, stimulus properties appear to be at least equal in importance to judge factors, if not strikingly more important, in determining the ratings of heterogeneous others, even when minimal information is given upon which ratings can be based.

Cross-Dimensional Consistency

Hypothesis II, that the tendency to intradimensional consistency on one scale is positively related to intradimensional consistency on all, was tested separately for the means and the variabilities of ratings made by the *Js* over all *Os* on all scales.

Consistency of means across scales. An analysis of variance was done in which each *J*'s unit scores were the means of his 30 ratings on each of the seven scales. The BJMS was then used to compute the r'''s of .16 for a single scale and .57 over all scales which appear for "All Scales" in Table 1. These r'''s, similar to those obtained for the individual scales, reveal only slight reliability for a single mean, but fair consistency across the total of seven means.

Further support for the hypothesis that consistency of means is stable across descriptive dimensions was found in a comparison of the confidence intervals of the seven intrascale coefficients of consistency. When the number of r'''s which failed to fall within the confidence interval of each coefficient was summed, it was found that in the case of none of the seven

TABLE 3
INTRACLASS CONSISTENCY COEFFICIENTS AND CONFIDENCE INTERVALS FOR MEANS AND VARIABILITY INDICES ACROSS SCALES

	Overall	Single Scale	Confidence Interval (Single Scale)
Means	.57	.16	.10–.32
Variability	.91	.58	.45–.69

scales did that number exceed the frequency that might have occurred by chance. The overlap among the confidence intervals attests to the fact that the seven r''s based on between judge differences for each scale are of approximately the same order of magnitude and, consequently, that consistency of mean response tends to be relatively constant across scales for the group of *J*s as a whole.

Consistency of variability across scales. Seven variability scores, one for each scale, were computed for each judge, representing his residual variability after the mean of his judgments of all *O*s, plus the mean group judgment of each *O*, respectively, plus the grand mean, were all partialed out of his ratings. The variability scores for the 62 *J*s were then ranked separately for each scale. An analysis of variance by ranks indicated highly significant between judge differences. The computed r''s were .58 for a single scale and .91 over all scales. The contrast between these r''s and the modest r''s for the consistency of the means is shown in Table 3, where the wide gap in the confidence intervals also can be seen. It will be noted that the r' for single scale variability is approximately equal to the r' for the total of seven means.

Sex Differences in Consistency

The investigation of the hypothesized "no difference" in consistency between the various subgroups of *J*s and *O*s consisted of applying the sign test to the frequencies with which the r' of one group exceeded that of the other on each of the seven scales. Since the widest split found between groups for any of the eight comparisons was 5:2, with a chance probability of .45, no significant trend seemed to be indicated. Thus no difference in consistency appears to distinguish the ratings of men *J*s from those of women, regardless of the sex of the *O*; neither do there appear to be differences in the consistency with which male and female persons are rated, regardless of the sex of the *J*.

DISCUSSION

Although marked individual differences in variability indicate that the data do not perfectly fit the assumptions of a parametric test, the results of this study, in conformity with those of Beier et al. (1957), Crow and Hammond (1957), and Izard (1959), reveal the presence of intrajudge consistency in impressions of heterogeneous persons.

However, the relatively slight power of observer bias is shown by the analysis of components of variance in which the proportion of variance found to be contributed by between judge differences is in line with the 3% found by DeSoto, Kuethe, and Wunderlich (1960). The frailty of individual bias is further paralleled in the findings of Campbell and Burwen (1956), whose reliability coefficients are of approximately the same order of magnitude as the intraclass correlations in the present study. Although the results do not explain the nonsignificance of respondent effects in three of Campbell and Burwen's four groups, they do suggest that, since the contribution of a general response set is small, any gross measure of this variable would be fragile, its appearance or nonappearance depending largely upon other factors accounting for more of the total variance. In contrast to the small influence of general response sets, the strength of specific attitudes is indicated by evidence such as Campbell and Burwen's finding of higher reliability coefficients for judgments of homogeneous as compared with heterogeneous groups of *O*s, and the positive findings of the numerous studies investigating specific determinants of social perception (e.g., Cohen, 1956; Scodel & Austrin, 1957; and Secord, 1958). It seems likely that the generalized *O* represents too high a degree of abstraction and that general response set is organized at too tentative a level of integration to exert a consistent influence upon behavior.

The highly significant stimulus effects found in this study have also been noted in previous research, e.g., by Beier et al. (1957), Campbell and Burwen (1956), and, most outstandingly of all, by Izard (1959), despite his care in

maximizing stimulus ambiguity. What can be the nature of this stimulus effect? Though no criterion data are available, it is improbable that the judgments had much validity. It seems much more likely that cultural similarities in the experience of the *J*s resulted in their drawing upon commonly held stereotypes. Indeed, the relative proportion of stimulus to respondent effects shown in the analysis of components of variance shows that, at least under conditions of low to moderate ego involvement, the *J*s relied more upon these commonly held sets than upon unique ones. The ratio of stimulus to respondent effects may possibly have been exaggerated by homogeneity among the *J*s, resulting not only in reduced r''s but also in increased agreement upon the ratings of the stimulus persons. Still, while heterogeneity of age, education, social class, etc. would shrink interjudge agreement, any resultant reduction in stimulus effects would be likely to accrue not to the judge factor, but to random variation. These results seem in conflict with the common assumption that the presentation of ambiguous stimuli evokes largely autistic elements in perception. They thus accentuate the importance of normative data in the interpretation of projective tests.

Although modest but statistically significant intrajudge consistency of means was found across scales, even larger between scale differences indicate that the group of *J*s tended to use the scales in the same way. Scale effects probably cannot be attributed solely to social desirability since the correlations found between the rank order of the scale means and Cowen, Budin, and Budin's (in press) normative values for "Personal" and "Social" desirability response sets toward the scale anchor terms were nonsignificant. ($N = 6$, no norms being available for the Masculine-Feminine vs. Effeminate-Mannish scale.) Another possible source of between scale differences may have been the intrajudge consistency of covariance patterns which Cronbach (1955) proposed as the third parameter of an "implicit personality theory."

The high reliability of individual variability appears to call for considerable further investigation. Although the phenomenon may be a manifestation of personal style in the use of the instrument rather than of the *J*'s characteristic range of social judgments, Broen and Wirt's (1958) correlation of .03 between two measures of the tendency to spread ratings on rating scales indicates that style of responding on rating scales, per se, is unlikely to have been a major factor in the consistency of variability scores. From the greater stability of these indices compared with the mean ratings, one might further speculate whether impulsivity, or expansiveness in judgments of others, may not be a more salient response set than favorability.

In sum, these findings cast doubt on theoretical positions like those of Kelly (1955), Sullivan (1956), and Cronbach (1955), according to which the judge's unique system of constructs, personifications, or implicit personality theory are central in determining his impressions of people. The modification of expectancy theories to accord with the results of this investigation would require a minimizing of the emphasis upon people's outlook on others in general and, instead, increased attention: (*a*) to the pull of aspects of the stimulus regarding which there is a large measure of interjudge agreement, and (*b*) to the role of individual differences in range of response. Finally, although less clear-cut on this last point, the results of this study, along with those of others, suggest that the theoretical concept of a unitary attitude toward the generalized other may be too loose to have more than severely limited significance and that such terms as "implicit personality theory" might with greater precision be used in the plural rather than in the singular.

SUMMARY

The existence and relative power of a unitary attitude toward diverse persons, implied in statements of Kelly (1955) and Sullivan (1956), were investigated by utilizing Cronbach's (1955) operational definition of "implicit personality theory." Although the results supported the concept of a general response set, the fragility of intrajudge consistency relative to stimulus factors points to a construct difficult to validate. No differences in consistency were found to be related to sex of *J* or *O*. The findings suggest less theoretical emphasis upon *J* bias and greater attention to response variability and "reality" as evinced

by interjudge agreement concerning stimulus attributes.

REFERENCES

ALFERT, ELIZABETH. Two components of assumed similarity. *J. abnorm. soc. Psychol.*, 1958, **56**, 135–138.

BEIER, E. G., SMOCK, C. D., IZARD, C. E., & TOUGAS, R. R. "Response to the human face as a standard stimulus": A re-examination. *J. consult. Psychol.*, 1957, **21**, 165–170.

BROEN, W. E. J., & WIRT, R. D. Varieties of response set. *J. consult. Psychol.*, 1958, **22**, 237–240.

CAMPBELL, D. T., & BURWEN, L. S. Trait judgments from photographs as a projective device. *J. clin. Psychol.*, 1956, **12**, 215–221.

COHEN, A. R. Experimental effects of ego-defense preference on interpersonal relations. *J. abnorm. soc. Psychol.*, 1956, **52**, 19–27.

COWEN, E. L., BUDIN, W., & BUDIN, FLORENCE A. The social desirability of trait descriptive terms: A variation in instructional set. *J. soc. Psychol.*, in press.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CROW, W. J., & HAMMOND, K. R. Generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

DESOTO, C., KUETHE, J. W., & WUNDERLICH, R. Social perception and self-perception of high and low authoritarians. *J. soc. Psychol.*, 1960, **52**, 149–155.

GUILFORD, J. P., & ZIMMERMAN, W. S. *Guilford-Zimmerman Temperament Survey manual.* Beverly Hills, Calif.: Sheridan Supply, 1949.

HAGGARD, E. A. *Intraclass correlation and the analysis of variance.* New York: Dryden, 1958.

IZARD, C. E. Paranoid schizophrenic and normal subjects' perceptions of photographs of human faces. *J. consult. Psychol.*, 1959, **23**, 119–124.

KELLY, G. A. *The psychology of personal constructs.* Vol. I. New York: Norton, 1955.

SCODEL, A., & AUSTRIN, H. The perception of Jewish photographs by non-Jews and Jews. *J. abnorm. soc. Psychol.*, 1957, **54**, 278–280.

SECORD, P. F. Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 300–315.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SULLIVAN, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1956.

(Received May 27, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 611-615

AN EXPERIMENTAL TEST OF SCHEMA INTERACTION[1]

DONALD P. SPENCE[2]
Research Center for Mental Health, New York University

WHEN a person is asked to describe a strange face, he notes the obvious physical features and from them makes assumptions about the person's possible feelings and other characteristics. These assumptions are partly implicit, based on past associations of physical features with specific experiences. Murray (1933) has described the working of these associations thus: "Certain physical features of the object which the subject does not consciously perceive are nevertheless physically affecting his body, and though he may be unable to report upon these internal happenings, they are nevertheless affecting his conscious appraisal of the object" (p. 312).

Thus it would appear that the description of a strange face may be partly determined by stimulation outside of awareness. In slightly different language, it seems that the meaning attributed to a given facial expression depends on which schemata are influencing perception, where a schema, following Bartlett (1932), is a semipermanent aggregate of connotations not necessarily in awareness. Klein and Holt (1960) have suggested that an incidental stimulus can activate such a schema; hence the stimulus should be able to change the expression attributed to a strange face. Smith, Spence, and Klein (1959) showed that such is the case; when the words happy and angry were presented below threshold, they influenced the description of a neutral face. Eagle (1959) showed that a subliminal aggressive picture can influence the attribution of aggressive traits to a neutral figure. Evidently the schema activated by the incidental stimulus interacts with the schema aroused by the face and results in a change in description. The more expereince the subject has had with the incidental stimulus, the more established is the associated schema, because schemata are built up from experience (Bartlett, 1932).

How do we learn about the schema activated by the incidental stimulus? By using the subject's description of the stimulus in full awareness. Previous experiments with subliminal stimulation (Eagle, 1959; Smith et al., 1959) have indicated that the stimulus that registers below awareness corresponds in large degree with the description of the stimulus in full awareness. Thus a rating of the stimulus should be a good approximation of the schema it activates when presented incidentally.

In the present study, the subject is asked to rate three emotional objects. This rating is supposed to crystallize a set of feelings about the objects—i.e., reinforce the object schema. In a later session, each object is presented incidentally, immediately prior to an ambiguous face. The study tests the hypothesis that the incidental object reactivates the schema mobilized by the initial rating, and that this schema then interacts with the face schema to bring about changes in description, causing the face to take on some of the properties of the object.

METHOD

Design and Procedure

The experiment consisted of two sessions averaging 90 minutes each and separated by an average interval of 2 months. In the first session the subject was asked to describe three pictures of faces which subsequently became the focal stimuli. The faces were relatively expressionless and yet easily discriminable. Each was exposed in a three-field tachistoscope for 15 seconds at an illumination of .8 footcandles, with the following instructions:

> I'm going to show you a face for a short look. Study it carefully. When I turn it off, I want you to lean back in the chair and tell me whatever comes to mind—what this face reminds you of, how it makes you feel, anything you like. Keep talking until I ask you to stop.

Associations to the face were recorded for 2 minutes. Then after another, 5-second exposure of the face, the subject described it on a rating form of 41 items. Each item was a dichotomy (e.g., "modest-boastful") expanded to a seven-point scale which provided three

[1] The study was carried out as part of a research project supported by grants from the Ford Foundation and the United States Public Health Service.

[2] The writer is indebted to George S. Klein for his help in setting up the experimental design, testing subjects, and for analysis and interpretation of the findings; to Robert R. Holt and Fred Pine for their suggestions and criticism; and to Paul Lippmann for his assistance in testing subjects and analyzing data.

degrees of choice on each side of a middle, "can't-tell" position. Items were chosen from Osgood (1958) to cover the three main factors of Evaluation, Potency, and Activity.

After the subjects had given associations and ratings to all three faces, he was asked to leave the room temporarily and in his absence 11 stimulus objects were arranged on a table. The objects were selected to elicit a wide range of feelings and consisted of pipe, baby bottle, earring, lipstick, female stationery, pistol, whiskey bottle, sympathy card, exam bluebook, ink bottle, charred stick. Then the subject was brought back and given the following instructions:

> Now I want you to look at the objects here on the table. Look them over carefully and choose one. When you have picked it, I want you to sit back, study it and think about it—what it is, and what it reminds you of. Then I want you to lean back, close your eyes and talk about whatever comes to mind—what the object makes you think of, and how it makes you feel. Keep talking until I say stop.

Order of selection was recorded for each subject. After he had given associations to the first object, it was shown to him for 5 seconds in the tachistoscope. He then rated it on the Osgood checklist. The same procedure was repeated for the remaining objects.

The associations and checklist ratings were used to choose the two pre-face incidental stimuli given to each subject. There were three criteria for selection: checklist profile with a high number of extreme ratings, free associations that contained many feelings and personal references, and extreme positive or negative connotations in the associations. If the stimulus was high on all three criteria, it was assumed to have a particularly polarized schema that would have a substantial effect on the face schema. The two most extreme objects for each subject became his later incidental stimuli; one was positive (e.g., baby bottle) and the other was negative (e. g., pistol).

The initial rating and free association to the faces and objects had the function of reinforcing one particular set of associations for each stimulus. Of the many possible schemata activated by each face or object, one was made relatively explicit by being recorded on a rating scale. When the face or object was shown in a later session, it was assumed that this particular schema would be activated.

The second session constituted the main part of the experiment. One to 3 months after the first session, each subject was re-exposed to the three faces, each face immediately preceded by one of the incidental stimuli previously chosen. Each face was exposed at 330 milliseconds and the incidental object at less than 30 milliseconds. It had been found in a previous study (Klein, Spence, Holt, & Gourevitch, 1958) that such a combination of speeds prevents the first, briefly exposed object from appearing in conscious awareness. To determine the exposure level for the pre-face stimulus, we established the subject's recognition threshold for a line drawing of a diamond, and then presented the stimulus object a certain level above the diamond threshold. The increment was determined by a small pilot study.

Exposures of the objects varied from 6 to 28 milliseconds. Median exposures were 11 milliseconds for the positive object, 12 milliseconds for the negative object, and 11 milliseconds for the neutral object. Since no subject reported seeing the pre-face object, our exposures appear to have been adequately incidental (in fact, probably subliminal most of the time). We may assume that the subject did not experience the objects as relevant and focal stimuli in his dealings with the face stimulus.

Each subject was exposed to the three faces in a constant order. The object-face pairing was counterbalanced with roughly half the subjects getting Face 1 with the positive object and Face 2 with the negative object. All subjects were given Face 3 with the neutral object (the charred stick). The object was mounted on pegboard, slipped into channels placed at either side of one exposure field of a three-field tachistoscope. The face was placed in the second field. The subject was asked to fixate on the third, or blank field; at a given signal the object was briefly exposed followed immediately by the face at 330 milliseconds. An adjustable electronic circuit controlled the rapid pairing of the stimuli, details of which may be found in Klein et al. (1958).

The subject was told that he would see a face, one he had seen before, at four brief exposures. After the exposure of Face 1 and its pre-face object, the subject was asked to associate to it for 2 minutes. After the association, face and object were presented two more times and the subject was asked to describe the face on the Osgood checklist. Faces 2 and 3 were presented with the same instructions. At the end of the session, the subject was asked to list all objects he could remember from the first session. Order of recall was recorded.

Subjects

Subjects were paid volunteers, 13 men and 10 women, who were undergraduates in a school of education. All subjects were part of a larger pool intensively studied on a variety of personality measures (see Goldberger & Holt, 1958, for details).

RESULTS

Were four brief presentations of the stimulus sufficient to activate the object schema and influence the rating of the face? It was expected that the object schema would interact with the face schema in such a way that the final rating would be a combination of the two earlier ratings. If we assume that the object was first rated "very strong" and the face was first rated "somewhat weak," a combination of these two ratings would make the face, on the second rating, move toward the "strong" end of the scale. Thus the combination of object and face schemata would result in a shift toward the original object rating. We tallied

all ratings that changed two or more units on the checklist from the first rating of the face to the second, and subtracted shifts *toward* the object from shifts *away* from the object. By chance, both types of shifts are possible; by hypothesis, only shifts *toward* the object are expected. Therefore, shifts toward the object should exceed shifts away. Data were tallied separately for each object, each subject being given a score according to whether he had more shifts away from the object or toward the object.

The majority of shifts were toward the object in all three object-face conditions, and the difference was significant with two. When paired with the negative object, the ratings of the face shifted toward the earlier ratings of the object; 15 out of 23 subjects ($p < .05$, one-tailed sign test) showed a preponderance of shifts toward the object. When paired with the neutral object, the face's ratings moved predominantly toward the object's ratings in 14 out of 23 cases ($p < .05$, one-tailed sign test). When paired with the positive object, the ratings of the face moved toward the ratings of the object in 12 out of 23 cases (*ns*) but 7 of these cases were women; thus the women behaved as predicted, and the interaction of sex with shift toward the object approaches significance ($p < .10$, Fisher exact test, two-tailed). In the main, therefore, the hypothesis was confirmed.

Additional Effects

When the explicit properties of the incidental stimulus (e.g., "very modest," "slightly healthy," etc.) are not taken into account and we consider only the general evaluative dimension, its influence on the face is much less predictable. Ratings of the face were scored on a 21-item evaluative scale, i.e., items were considered as positive or negative, according to a consensus of four judges. When the objects were scored on this scale they were markedly different, with the positive objects having a mean rating of 18 and the negative objects a mean rating of 5. Thus the positive objects were positively evaluated and the negative objects the reverse. When the final ratings of the faces are scored on this scale, we find they were not influenced by the global properties of the objects; on the contrary, the faces are not significantly more positive after being paired with the positive object nor significantly more negative after being paired with the negative object. The specific properties of the object schema, rather than general "goodness" or "badness," seem thus to have been transferred to the faces in the interaction of object and face schema.

Memory

Memory of the *neutral* incidental object was significantly related to its effect on the associated face. At the end of the object-face pairings, subjects were asked to list all the objects they could remember from the earlier session. Some mentioned the neutral object early on the list, some later, and some not at all; subjects were ranked according to ease of recall, with a low rank indicating early recall and a high rank indicating late recall or forgetting.

Subjects were ranked for degree of incidental effect by summing the algebraic total of shifts away from and toward the neutral object. The subject with the largest net number of shifts toward the object was given a rank of 1.

Order of memory of the neutral object correlated .56 ($p < .02$, two-tailed test) with degree of effect; subjects whose impresions of the face showed most direct influence by the object were also most likely to remember the neutral object.

Memory of the *positive* and *negative* objects was not significantly related to amount of effect.

DISCUSSION

The hypothesis that an incidental stimulus activates schemata and thus can influence the description of a face has been supported by the negative and neutral stimulus objects, and by the positive object for female subjects. The schema activated by the incidental object appears to combine with the face schema in a predictable fashion and to influence the description of the face. The combination of schemata seems to follow fairly straightforward additive principles, with one attribute increasing or decreasing the strength of a related attribute: e.g., "active" added to "passive" shifts the balance toward "active."

The results are in line with those of earlier

studies by Smith, Spence, and Klein (1959) and Bach and Klein (1957) in which subliminal stimuli appear to influence descriptions of supraliminal figures. In these studies a verbal stimulus outside of awareness (happy or angry) apparently activated a congruent schema and when the subject was asked to describe the supraliminal figure his description was influenced by the activated schema. The present study differs from these two in showing that the discrete attributes of the activated schema may interact with attributes of the face schema, thus supporting the assumption that an incidental stimulus can be faithfully registered outside of awareness in more or less specific detail, and that a generalization of effect does not necessarily take place. Thus the negative object did not make the face appear generally more negative, nor the positive object make it more positive; on the contrary, the effect was tied to specific object properties which represented a mixture of goodness and badness.

When the stimulus is sufficiently compelling, a generalized incidental effect may occur. Eagle (1959) has shown that a subliminal exposure of a man being stabbed makes for a general increase in aggression attributed to the following stimulus; Pine (1960) showed that an oral-passive incidental stimulus makes for stories characterized by an oral-passive theme, but not necessarily containing specific elements of the stimulus. In both cases, the incidental stimulus was heavily loaded with drive content and, through contact with instinctual wishes, its effect may be increased. Drive energy is presumed to follow the rules of the primary process—hence it is less bound by realistic notions of form and content and is more easily displaceable. Under such conditions, it seems reasonable that an incidental drive stimulus would have less precise effects. An incidental stimulus that made less contact with drive material would result in a more veridical effect, as was the case in the present study. It is interesting, in this connection, that the positive object, in some ways the most drive related because of its strong oral-passive connotations, showed the least veridical effect.

It should be noted here that the more veridical effects of the present study are easier to trace because the prior object rating provides a baseline. In Pine (1960) and Eagle (1959), there was no baseline and the stimulus that registered may have differed from the objective stimulus presented. In these studies the response may have been closely related to the *registering* stimulus and not, strictly speaking, a transformation. In the present study, the rating profile enables us to get closer to the registering stimulus and cuts down one possible source of error. Thus the greater veridical effect may be partly a consequence of method as well as the result of less drive.

Congruence

The failure of the positive object to conform to the hypothesis may stem, as already noted, from its greater drive relevance, but another possibility should be considered. All positive objects (e.g., baby bottle, female stationery) had (inadvertently) strong feminine connotations and they were paired with masculine faces. Now the connotations of any attribute used to describe a feminine object are quite different from its connotations when it is used to describe a male object. "Passive," for example, may endow a feminine object with pleasant and favorable qualities, but when applied to a masculine face it has other implications. Passive obviously has different connotations in the two contexts and it would be misleading to assume that it has the same meaning in both contexts. Yet that is the assumption behind the experimental hypothesis—that passive has identical meanings in all contexts, and attributes can be added and subtracted like so many counters.

This example suggests a general rule: the way in which two schemata interact depends on their overall contexts. If two schemata have roughly equivalent contexts, attributes from Schema I should have roughly the same connotations as attributes from Schema II, and the two sets of attributes can be treated as equivalent and combined without loss of meaning. What Eagle (1959) calls congruence seems to be an important variable in predicting the interaction of stimuli. If the congruence is low, the contexts are different, and the attributes of the separate schemata have different connotations; as a result, the final rating may

shift away and toward the object and the final effect should be difficult to predict.

In general, we should expect to find more additive effect when congruence is high, and the results bear out this expectation. The negative and neutral objects that produced the strongest direct effect were either masculine in content or in form—e.g., pistol, whisky bottle, square sympathy card, and angular stick—and thus more congruent with the masculine faces. The positive objects were more feminine, less congruent, and less effective.

SUMMARY

In an attempt to test propositions about the interaction of schemata, three stimuli, incidental to the stimulus to be rated, were paired with three supraliminal faces in a partly balanced design. Twenty-three subjects rated the stimuli and faces initially on an Osgood Semantic Differential. In a second session, several months later, they rated each face again, preceded by one of the objects below the level of focal awareness. It was predicted that the object and face schemata aroused by the first rating would be reactivated, and would interact in a predictable way to produce a new rating of the face. For two of the three types of objects, ratings of the face following the negative and control incidental stimuli moved in the direction of the object rating, in support of the hypothesis. The findings were discussed in terms of a theory of incidental schemata that influence ongoing cognition.

REFERENCES

BACH, S., & KLEIN, G. S. The effects of prolonged subliminal exposures of words. *Amer. Psychologist*, 1957, **12**, 397–398. (Abstract)

BARTLETT, F. C. *Remembering*. Cambridge, England: Cambridge Univer. Press, 1932.

EAGLE, M. The effects of subliminal stimuli of aggressive content upon conscious cognition. *J. Pers.*, 1959, **27**, 578–600.

GOLDBERGER, L., & HOLT, R. R. Experimental interference with reality contact (perceptual isolation): Method and group results. *J. nerv. ment. Dis.*, 1958, **127**, 99–112.

KLEIN, G. S., & HOLT, R. R. Problems and issues in current studies of subliminal activation. In J. G. Peatman & E. L. Hartley (Eds.), *Festschrift for Gardner Murphy*. New York: Harper, 1960. Pp. 75–93.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

MURRAY, H. A. The effect of fear upon estimates of the maliciousness of other personalities. *J. soc. Psychol.*, 1933, **4**, 310–329.

OSGOOD, C. *The measurement of meaning*. Champaign, Ill.: Univer. Illinois Press, 1958.

PINE, F. Incidental stimulation: A study of preconscious transformation. *J. abnorm. soc. Psychol.*, 1960, **60**, 68–75.

SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

(Received June 3, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 616–622

ATTITUDES TOWARD OLD PEOPLE IN AN OLDER SAMPLE[1]

NATHAN KOGAN[2]
Educational Testing Service

In a recent paper (Kogan, 1961), the present author described the development of a scale measuring attitudes toward old people (OP scale). For college samples, a positively and a negatively worded version of the OP scale were significantly correlated in the content direction, and scale reliabilities were generally satisfactory. In addition, significant relationships that could not be attributed to response set effects were found between OP scale scores and other attitudinal and personality variables.

This paper reports results obtained from the administration of the OP scale to a sample of apparently healthy older adults of better-than-average intelligence and education. With college students' responses serving as "baseline" data, OP scale characteristics are examined in the older sample with particular reference to the favorability and variability of the attitude, the internal consistency of the scales, the logical consistency in response to positively and negatively worded versions of the scales, and direction and magnitude of relations with other attitudinal variables.

Specific predictions of differences between the younger and older samples were derived from two general assumptions. First, it was assumed that older people in American society have many of the qualities of a minority group, a point that has been stressed by a number of investigators (Barron, 1953; Drake, 1958; Linden, 1957a, 1957b). If Lewin's (1945) concept of "self-hatred" is applied in the present case, the empirical consequences might well take the form of a greater polarization in the attitude; older people would be more likely than younger respondents to be found at either extreme regarding the issue. Hence, older adults should manifest greater inter-individual variability. Second, it was assumed that there would be a greater "ego involvement" on the part of older people in the content of the OP scale items. As a member of a group that comprised the object of the attitude under study, an older person might be expected to take a closer, more considered look at each OP item than would a younger one. Accordingly, we should expect older subjects to exhibit somewhat greater logical consistency in responding to matched positive and negative OP items.

No hypotheses were formulated with respect to differences between younger and older subjects in the favorability level of attitudes toward old people. If older individuals exhibit considerable variability in the extent of acceptance-disparagement of old people as a group, it is difficult to predict whether older subjects should be more or less biased than younger subjects.

Similarly, no hypotheses are advanced respecting the magnitude of relations between OP scale responses and responses to other attitudinal content in the older and younger samples. It will first be necessary to examine the extent to which stylistic response tendencies can account for relations between the variables.

METHOD

Subjects

Older subjects (89 males and 115 females) were obtained from the Age Center of New England, Inc., a nonprofit gerontological research organization in Boston, Massachusetts. These subjects were volunteers, noninstitutionalized, and in apparently good health, whose membership in the organization involved a commitment to participate in a variety of research studies. Chronological ages of the subjects ranged from 54 to 92 for the males, with a mean of 70.8 (SD = 7.3), and from 49 to 86 for the females, with a mean of 69.4 (SD = 7.9).

Approximately 70% of the older subjects held or

[1] This study was supported in part by the National Institute of Mental Health, United States Public Health Service, under Research Grant M-1867, and was conducted under the auspices of the Age Center of New England, Inc. Support for machine processing of much of the data was provided by Educational Testing Service.

[2] Grateful thanks are due Samuel J. Messick for statistical advice and critical comments on the paper. The author also wishes to express his appreciation to Beverly Scott, Florence C. Shelton, and Judith W. Stephens for their assistance in data collection and analysis.

are now holding professional or managerial positions; most of the remaining 30% belong to the clerical or skilled craftsman occupational groups. The mean educational level in the sample was 13.7 (SD = 2.8) and 12.9 (SD = 3.1) years of school for males and females, respectively. *Current Population Reports* (Bureau of the Census, 1960), citing years of school completed for the civilian population of the United States in March 1959, reports a median figure of 8.7 for males 55–64 years and 8.2 for males 65 years and over. The corresponding figures for females of those ages are 8.9 and 8.4 years of school, respectively. By contrast, median school years completed is 12.4 for both males and females in the 20–24 age group. Hence, the educational level of the older subjects could well be considered superior to a contemporary college student group. This inference is further supported by intelligence data available on a random subsample of 52 males and 73 females, who were given the vocabulary subtest of Wechsler's (1944) intelligence test for adults. The mean scores for males and females were 34.7 (SD = 4.8) and 33.9 (SD = 5.4), respectively. Both means are slightly larger than those reported by Dana (1957) for college student samples.

Young adult subjects (87 males and 81 females) were obtained from introductory classes in social science at Boston University.[3]

Measuring Instruments

A set of 34 "old people" items, 17 expressing negative sentiments about old people (OP−) and 17 reversed in wording expressing positive sentiments (OP+), were interspersed among statements from other attitude domains yielding a 98-item battery for the older sample and a 72-item battery for the student subjects. Common to both samples, in addition to the OP scales, were 12-item F and reversed F scales and 6-item "religious conventionalism" (RC) and reversed RC scales. The particular items used in the OP and F scales are reported in Kogan (1961). RC items were selected from Levinson (1954) and a set of matched items of opposite sign was constructed (Table 1).

The following response categories were provided for all of the items: "strongly disagree," "disagree," "slightly disagree," "slightly agree," "agree," and "strongly agree." These were scored 1, 2, 3, 5, 6, and 7, respectively, with a score of 4 assigned in the few cases where respondents omitted an item.

RESULTS

Mean scale scores for young and old subjects are shown in Table 2. While no age differences emerge for the OP− scale, there is a highly significant difference for OP+, with older subjects exhibiting a greater tendency to endorse such items. The latter finding when viewed relative to the direction of the mean differences for OP− would suggest that an acquiescent response set exerted a stronger influence in the older sample.

[3] I am grateful to Robert Smart for making subjects available.

TABLE 1
ORIGINAL AND REVERSED RC SCALE ITEMS

1(+).	Every explanation of man and the world is incomplete unless it takes into account God's will.
1(−).	God's will should play no part in explaining man and the world; scientific knowledge is all that is needed.
2(+).	The chief end of man is nothing other than eternal salvation.
2(−).	The promotion of individual personal happiness is the chief end of man.
3(+).	Every person should have complete faith in some supernatural power whose decisions he obeys without question.
3(−).	There is no faith or supernatural power whose commandments one should obey without question.
4(+).	In addition to faith we need help from God in order to resist temptation.
4(−).	Prayer and faith in God don't help very much in resisting temptation.
5(+).	The most important aim of the churches at the present time should be to encourage spiritual worship and a sense of communion with the highest.
5(−).	The most important aim of the churches at the present time should be to encourage man's humanity to man.
6(+).	Life could hardly be worth living without the promise of immortality and life after death.
6(−).	The promise of immortality and life after death is quite unnecessary for a worthwhile and happy life.

Such an interpretation, however, is not unequivocally supported by the mean differences observed for the F and RC scales. With regard to authoritarianism items, it is the older men who score significantly higher in authoritarianism on F+, while the older women are significantly lower in authoritarianism on F−. While both sets of differences can be attributed to a stronger acquiescent response tendency in the older subjects, the fact remains that such tendencies were not of equal strength for original and reversed items. Similar findings may be noted in the domain of religious conventionalism. The older males achieve significantly higher scores on RC+ items, with no difference obtaining for RC−. The older women, on the other hand, manifest more religious conventionalism on both RC+ and RC− items, a finding directly counter to a simple acquiescence interpretation. In sum, it is evident that mean scale differences between samples reflect a combination of set and content, with the relative contribution of set related to the particular content area of a scale.

It will be recalled that we hypothesized

TABLE 2
MEAN DIFFERENCES BETWEEN YOUNG AND OLD SUBJECTS ON OP, F, AND RC SCALES

Scale	Young Men (N = 87)		Old Men (N = 89)			Young Women (N = 81)		Old Women (N = 115)		
	M	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
OP−	55.85 (3.28)[a]	11.73	57.26 (3.36)	11.97	.79	53.78 (3.16)	10.21[b]	55.83 (3.28)	12.61[b]	1.26
OP+	64.06 (3.77)	11.10	56.91 (3.34)	9.84	4.52**	64.30 (3.78)	10.73	57.79 (3.40)	11.37	4.08**
F+	42.24 (3.52)	8.44[c]	51.84 (4.31)	12.09[c]	6.12**	47.29 (3.94)	9.09[c]	49.31 (4.11)	12.29[c]	1.32
F−	46.04 (3.84)	6.33[b]	44.83 (3.74)	8.01[b]	1.11	48.60 (4.05)	6.63	42.84 (3.57)	7.30	5.74**
RC+	23.80 (3.96)	8.50	29.45 (4.91)	8.18	4.49**	24.62 (4.10)	7.64	29.96 (4.99)	9.01	4.47**
RC−	26.83 (4.47)	6.75	26.90 (4.48)	6.06	.07	25.72 (4.28)	6.00	27.94 (4.65)	5.92	2.57*

Note.—Converted means are shown for the OP+, F−, and RC− scales in order that means for positive and negative scales be comparable, i.e., higher values reflecting more negative attitudes toward old people, more authoritarianism and more religious conventionalism.
[a] Per item means.
[b] Variances differ significantly at .05 level.
[c] Variances differ significantly at .01 level.
* Significant at the .05 level.
** Significant at the .01 level.

TABLE 3
ODD-EVEN RELIABILITY COEFFICIENTS (SPEARMAN-BROWN) AND INTERSCALE CORRELATIONS FOR YOUNG AND OLD SUBJECTS

	OP−		OP+		F+		F−		RC+		RC−	
OP−	.83	.79	.60**	−.02	.46**	.46**	.17	−.21*	.30**	.04	.05	−.36**
	.75	.84	.41**	.05	.16	.46**	.04	−.19*	.22*	.24*	.08	−.31**
OP+			.74	.73	.07	−.35**	.11	.34**	.16	−.21*	.13	.31**
			.79	.82	−.30**	−.35**	−.04	.17	.00	−.24*	.17	.10
F+					.61	.80	.48**	.15	.29**	.41**	.04	−.18
					.59	.81	.31**	.23*	.26*	.42**	.03	−.18*
F−							.23	.23	−.02	.04	−.03	.39**
							.25	.28	.08	−.13	.31**	.14
RC+									.86	.82	.76**	.42**
									.71	.90	.65**	.44**
RC−											.71	.63
											.51	.53

Note.—The four samples of 87 young men (YM), 81 young women (YW), 89 old men (OM), and 115 old women (OW) are arranged in the cells of the table as follows:

YM	OM
YW	OW

* Significant at the .05 level.
** Significant at the .01 level.

greater variability among older subjects in their attitudes toward old people. Only for females on the OP− scale were the variances significantly different in the predicted direction. Variances were significantly different between age groups, however, for both males and females on F+, and for males on F−. This may simply reflect the greater range of education in the older samples (Christie, Havel, & Seidenberg, 1958).

Reliability coefficients and intercorrelations between the various scales are shown in Table 3. With the exception of the F− scale, reliabilities for the four samples are of moderate to high magnitude. Turning to relationships between positively and negatively worded

versions of the three scales, it should be noted that *r*'s can theoretically range from +1.00 (perfect logical consistency) to −1.00 (perfect logical inconsistency). It is apparent that *r*'s are uniformly higher in the content direction for younger subjects. Indeed, with the exception of *r*'s for females on the F− scale, differences between the younger and older samples in positive-negative scale correlations are statistically significant beyond the .05 level or better (two-tailed test).[4] Accordingly, we must abandon the hypothesis that the greater involvement of older subjects in OP item content is reflected in a higher correlation between oppositely worded versions of the OP scale. Particularly puzzling is the fact that the absolute level of *r* between OP− and OP+ for older subjects is less than that observed for F and RC items. Both of the latter would seem to be of less direct relevance to older subjects than are items explicitly referring to various attributes of old people.

Consider next the relations between the OP, F, and RC scales reported in Table 3. In every case, significant (positive) *r*'s between scales based on items worded in the same direction become nonsignificant or significant in the set (negative) direction when *r*'s are computed between scales of opposite wording. In sum, it must be concluded that there are no clear content relationships between the three attitudinal domains under study for either young or old adults. An attempt to explain the absence of a relationship between OP and F in younger samples may be found in Kogan (1961).

DISCUSSION

The results reported above generally do not support the hypotheses advanced earlier in the paper. In but one case out of four did older subjects exhibit greater variability in their attitudes toward old people. Correlations between positively and negatively worded versions of the OP scale not only failed to exceed those reported for younger subjects, but actually were of approximately zero magnitude.

Do these results call for a modification of the assumptions from which hypotheses were derived? It will be recalled that the older sample was distinguished from the younger in terms of the presence of tendencies toward disparagement of one's age group and "ego involvement" in the issue at hand. These assumptions appear so very basic in terms of logic and previously reported evidence that one is prompted to look elsewhere in accounting for the unexpected results of the study. It is most likely that the difficulty lies in the nature of the transition from assumption to hypothesis. In particular, does the well-known phenomenon of assimilation of "majority group" values in some "minority group" members necessarily imply greater interindividual variability for the minority than the majority in responses to "minority group" items? Further, does stronger personal involvement in an issue necessarily imply a heightened awareness of the logical oppositeness of statements regarding that issue?

It is a generally accepted view (Kuhlen, 1959; Riegel, 1959) that the role of the older person in American society is one likely to give rise to feelings of denial and ambivalence regarding one's aging. A number of investigators (Kutner, Fanshel, Togo, & Langner, 1956; Phillips, 1957; Tuckman & Lorge, 1954) report that substantial numbers of persons in their sixties and above decline to classify themselves as "elderly" or "old." Given the inexorable physical changes accompanying aging and their social stimulus value, the attitudes of many older persons toward their age peers should be characterized by ambivalence. The operational consequent of such ambivalence may well be an increased intraindividual variability in this attitude domain. The comparability of scale reliabilities for young and old subjects indicates that such differential variability is not manifested within uniformly positive or negative item sets. Rather, it is expressed in the extent to which younger and

[4] Two-tailed tests were employed for OP scale comparisons for two reasons. First, a result in the direction opposite to that predicted could not be considered psychologically meaningless given the primitive state of theory in this area. Hence, the present problem does not seem to meet criteria proposed by Kimmel (1957) for the use of one-tailed tests. Second, the difference between the OP interscale *r*'s for young adults and maximum possible *r*'s for older subjects (the product of the square roots of the reliabilities) is considerably less than the difference between the former and the minimum *r*'s possible for older subjects. In other words, it is statistically more difficult for the OP interscale *r*'s of the older subjects to be larger rather than smaller than the corresponding *r*'s for younger subjects.

older subjects are consistent in responding to matched items of reversed wording.

Are we dealing here with anything more than differences in response sets in younger and older persons? There is little doubt in the present data that older subjects are more prone than younger subjects to "set" as opposed to "content" consistency in responding to attitude items. This does not imply, of course, that older individuals make more response errors when filling out written questionnaires. Rather, the phenomenon of acquiescence can probably best be considered an important personality correlate of aging or an aspect of an "old age" syndrome. This view is consistent with recent theory and research regarding response set as a stylistic personality variable in its own right (Couch & Keniston, 1960; Jackson & Messick, 1958). Given the energy loss, passivity and dependency alleged to be associated with old age, one might well expect older individuals to differ from their younger counterparts in willingness to endorse "authoritative" statements. Indeed, Jackson (1959) has recently demonstrated a relationship between response acquiescence and "cognitive energy" level.

The foregoing observations do not, however, account for the fact that the magnitude of the response set effect varied greatly across scales for older subjects and was ordered in a direction contrary to expectations, i.e., least content consistency on the OP scales. This finding suggests that the greater emotional involvement of old people in the attitude domain under consideration may have one of two alternative effects. Either the effect is one of a basic failure to apprehend the logical incompatibility in the paired statements, or, more likely, the feelings and experience of the older person are such that the matched statements may be logical, but not psychological, opposites. In the latter case, the inference is made that the attitude domain is less likely to be unidimensional for an older than a younger subject. More specifically, the older person's ambivalence regarding aging and his direct exposure to individual differences among age peers may be contrasted with the overgeneralized view, positive or negative, that college age subjects are likely to form of a group relatively remote from them in age status.

The older subject confronted with a statement describing "good" qualities of old people, and a matched statement describing the absence of such qualities or opposing "bad" characteristics, may conceive of the former as fitting a particular kind of old people, the latter as appropriate to a different kind. In most cases, of course, the respondent, if not dissociating himself completely from the class of old people, very probably considers himself as the sort of old person described in the positively worded items.

There is some evidence for this interpretation in the OP item statistics[5] for the younger and older subjects. The item yielding the highest absolute mean (most unfavorable response) for older subjects and the greatest mean discrepancy between younger and older subjects (younger subjects more favorable) reads as follows: "If old people expect to be liked, their first step is to try to get rid of their irritating faults." This evidence further serves to clarify the psychological basis for failing to endow old people with "logically" consistent qualities. For, while an older subject may have a generally positive view of old people, a sensitivity to "irritating faults" necessarily results in his agreeing with statements expressing those faults of special annoyance to him. Given, in addition, a stronger disposition to acquiesce in older persons, the observed differences between younger and older subjects assume a meaningful pattern.

The present work may be of relevance to the broader area of minority-majority group relations. It will be acknowledged that the criteria for belonging in the "minority group" of old people are considerably more ambiguous than those for ethnic or religious group membership. Nevertheless, a search for generalities might prove of value.

Prentice (1956) obtained very high correlations in the content direction between positively and negatively worded religious and

[5] A three-page table giving OP item means, standard deviations, item-sum correlations and matched item pair correlations for the young and old samples has been deposited with the American Documentation Institute. Order Document No. 6766 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

racial minority items. Would these high correlations be maintained for minority group members responding to matched items of reversed wording whose content concerns the minority in question? Would the magnitude of the correlation reflect the extent of "marginality" (Glaser, 1958) in the group? Of course, it would be necessary in studies of this kind to control for response set effects by including matched positive and negative items from other attitude domains. With response set controlled, it might prove possible to obtain item discrepancy scores for the individual minority group member, which would reflect major areas of ambivalence in one's ethnic or religious identification.

In a more practical vein, do the low OP interscale correlations for older subjects essentially rule out attitudes toward old people as a metric that can be studied in terms of its relations with other variables in older samples? This would appear to be an extreme view. While the OP interscale correlations hovered around zero for the older samples, *r*'s between specific item pairs exhibited considerable variation. For five of the OP item pairs, these *r*'s were statistically significant for both male and female older subjects. There are, in other words, certain aspects of the attitude on which there is more consensus and, correspondingly, less ambivalence among older persons. The addition of other items of this kind and deletion of item pairs correlating nonsignificantly could well raise the interscale *r* to the levels obtained for younger samples. Alternatively, one might consider dividing the OP scales into unidimensional subscales, each reflecting a particular aspect of the attitude domain. The major focus of the present paper, however, has been one of accounting for differences between older and younger subjects in responses to a common set of items.

SUMMARY

The present paper compares the responses of college students and noninstitutionalized older people of comparable education and intelligence to a set of statements measuring attitudes toward old people (OP scale), authoritarian tendencies (F Scale), and religious conventionalism (RC scale). For each scale, matched positively and negatively worded items were used. Hypotheses formulated for the OP scales—that older subjects would show greater interindividual variability and higher interscale correlations than younger subjects—were not confirmed. In fact, *r*'s between the OP+ and OP− scales for the older subjects were approximately zero, and significantly less than those obtained for the younger samples. In the older samples, interscale *r*'s for the F and RC scales, while higher in the content direction than the interscale *r*'s for OP, were nevertheless significantly less than the corresponding *r*'s obtained in the student samples. On the basis of these results (in conjunction with the pattern of young-old mean differences on the scales), it was apparent that older subjects are more disposed than younger individuals toward acquiescent response tendencies.

In accounting for age differences in OP scale results, emphasis was placed upon the ambivalence of the older person regarding his own age status and that of his age peers, with the result that logically incompatible statements were rendered psychologically consistent. By contrast, younger subjects were imputed to have a more consistently negative or positive view of "old people," with the consequence of a closer match between logical and psychological consistency.

REFERENCES

BARRON, M. L. Minority group characteristics of the aged in American society. *J. Geront.*, 1953, **8**, 477–482.

BUREAU OF THE CENSUS. *Current population reports, population characteristics.* (Series P-20, No. 99) Washington, D. C.: BC, 1960.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F Scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, **60**, 151–174.

DANA, R. H. A comparison of four verbal subtests on the Wechsler-Bellevue, Form I, and the WAIS. *J. clin. Psychol.*, 1957, **13**, 70–71.

DRAKE, J. T. *The aged in American society.* New York: Ronald, 1958.

GLASER, D. Dynamics of ethnic identification. *Amer. J. Sociol.*, 1958, **23**, 31–40.

JACKSON, D. N. Cognitive energy level, response acquiescence, and authoritarianism. *J. soc. Psychol.*, 1959, **49**, 65–69.

JACKSON, D. N., & MESSICK, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

KIMMEL, H. D. Three criteria for the use of one-tailed tests. *Psychol. Bull.*, 1957, **54**, 351–353.

KOGAN, N. Attitudes toward old people: The development of a scale and an examination of correlates. *J. abnorm. soc. Psychol.*, 1961, **62**, 44–54.

KUHLEN, R. G. Aging and life adjustment. In J. E. Birren (Ed.), *Handbook of aging and the individual.* Chicago: Univer. Chicago Press, 1959. Pp. 852–897.

KUTNER, B., FANSHEL, D., TOGO, ALICE M., & LANGNER, T. S. *Five hundred over sixty.* New York: Russell Sage Foundation, 1956.

LEVINSON, D. J. The intergroup relations workshop: Its psychological aims and effects. *J. Psychol.*, 1954, **38**, 103–126.

LEWIN, K. *Resolving social conflicts.* New York: Harper, 1945.

LINDEN, M. E. Effects of social attitudes on the mental health of the aging. *Geriatrics*, 1957, **12**, 109–114. (a)

LINDEN, M. E. Relationships between social attitudes toward aging and the delinquencies of youth. *Amer. J. Psychiat.*, 1957, **114**, 444–448. (b)

PHILLIPS, B. S. A role theory approach to adjustment in old age. *Amer. sociol. Rev.*, 1957, **22**, 212–217.

PRENTICE, N. M. The comparability of positive and negative items in scales of ethnic prejudice. *J. abnorm. soc. Psychol.*, 1956, **52**, 420–421.

RIEGEL, K. F. Personality theory and aging. In J. E. Birren (Ed.), *Handbook of aging and the individual.* Chicago: Univer. Chicago Press, 1959. Pp. 797–851.

TUCKMAN, J., & LORGE, I. Classification of the self as young, middle aged, or old. *Geriatrics*, 1954, **9**, 534–536.

WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

(Received June 6, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 623-630

A CROSS-LINGUISTIC INVESTIGATION OF PHONETIC SYMBOLISM[1]

MURRAY S. MIRON[2]
University of Illinois

A series of recent experiments have attempted to determine the universality of phonetic symbolism in languages. In a study by Brown, Black, and Horowitz (1955), subjects were given the task of choosing from a pair of antonyms, in a language familiar to them, that word which represented a translation equivalent from a second pair of antonyms in a language unfamiliar to them. A study by Maltzman, Morrisett, and Brooks (1956), in which a single stimulus word was to be matched with one of a pair of response words, also employed materials in languages unfamiliar to the subjects in both the stimulus and response positions. Still another variant procedure was employed by Brackbill and Little (1957). In this study subjects were required to make same-different judgments for pairs of single words; the members of the pairs being from different languages, one or both of which were unfamiliar to the subjects. All procedures, however, have aimed at demonstrating greater than chance accuracy in the choice of the unknown translation equivalent. This is not the place for a lengthy statement of the difficulties inherent in attempting to establish a workable criterion of translation equivalence, except to point out that linguists have generally despaired of any really satisfactory, analytic criterion. Brackbill and Little (1957) have pointed out that the selection procedures for the items used by these investigators have not been sufficiently rigorous. It is doubtful whether any selection procedure, no matter how analytic, can avoid the possibility of bias. A translator faced with the problem of choosing a translation equivalent from among several alternative words of varying meaning overlap might well condition his choice on the basis of a similarity in sound or form. If the translations are as dissimilar as possible, no phonetic symbolism effect would be expected. It is also clear that not all words in any given language are faithful examples of phonetic symbolism. It seems apparent, therefore, that the demonstration of any such effect may be crucially conditioned in a biased fashion by precisely the same variables that are under investigation.

Perhaps a better beginning point for demonstration of universality of phonetic symbolism would be one which attempts to determine the actual meanings *assigned* to various phonetic elements in nonsense combinations by subjects from diverse linguistic communities. Similarity of assigned meanings for subjects using different languages, although not directly attacking the problem of phonetic symbolism in natural languages, at least would assure us that some basis for a linguistic symbolic communality is tenable. Accordingly, the purpose of this study is to determine: (*a*) if systematically sampled speech sounds arranged in nonsense combinations display consistently different meanings when judged on certain descriptive (affective) dimensions; (*b*) if the changes in meaning bear any consistent, lawful relation to the phonetic properties of the sounds; and (*c*) if the meaningful differentiations of speech sounds and their relations to phonetic properties correspond across two contrasting linguistic communities, American English versus Japanese.

METHOD

Subjects

The American English-speaking subjects were 79 undergraduate and graduate students of both sexes in residence at the University of Illinois. All subjects in this group spoke English as a native language and did not have any knowledge of Japanese. The Japanese-speaking subjects were 41 Japanese foreign students and staff members with varying competence in English, employed by or attending the University of Illinois. All subjects in this group spoke the Tokyo dialect of Japanese and had resided in this country an average of 13.9 months at the time of testing. Although all Japanese subjects were familiar with English, a total of 17

[1] This paper is based on a doctoral dissertation presented to the University of Illinois.

[2] The writer gratefully acknowledges the supervision and guidance of Charles E. Osgood.

of the 41 subjects in this group had entered the country less than 3 months before testing. A single subject had resided in the United States for a total of 54 months and represented the maximum length of time in the country for this group.

Stimulus Materials

After a careful search of several languages historically unrelated to English for correspondence in the phonemic inventory, Japanese was finally selected. The phones selected are displayed in Table 1 with the specific combinations employed in the experiment. The choice of these combinations was conditioned by three criteria: meaninglessness of the stimulus in both English and Japanese, equal representation of each of the sound elements in the total list with equal number of occurrences of each consonant element before and after each of the five vowel elements, and adherence to the phonemic distribution laws of each of the languages. Preliminary work with Japanese bilingual informants uncovered six stimulus items that were identified as real words in Japanese. An identical number were found to be listed in English source books. Exclusion of these real words found in English and Japanese was made impossible by the second and third criteria listed above.

The final stimulus list was recorded on high quality equipment as spoken by a trained phonetician.[3] Because of the presence of minor differences between the articulation of the vowel and consonant phones in the two languages, a series of compromises in the production of the sounds was attempted by the speaker. These compromises were designed to favor neither of the groups with respect to the articulations. After recording the items the experimenter and the phonetician selected the one item of several attempts on each item which best approximated the desired effect both with respect to articulation and technical recording criteria. These items were then cut out of the original tape and spliced end-to-end to form a continuous loop that could be played repetitively.

Measuring Instrument

In order to index the symbolic meaning of the stimulus materials, a 15-scale form of the semantic differential was constructed. The scales were chosen to represent, in part, the factors of evaluation, potency, and activity previously found in factor analyses of various stimuli (see Osgood, Suci, & Tannenbaum, 1957) and in part to reflect various sorts of physical dimensions deemed appropriate to the evaluation of sounds.

The scales so chosen were combined, for purposes of analysis, into three composite scales identified as an evaluative composite (E), a potency composite (P), and an activity composite (A). The scales chosen to comprise the evaluative composite were: good-bad, beautiful-ugly, pleasant-unpleasant, high-low, and colorful-colorless. Those comprising the potency composite were: heavy-light, powerful-powerless, thick-thin, large-small, hard-soft, and strong-weak. The activity composite was derived from the scales: quiet-noisy, slow-fast, passive-active, and cold-hot. One additional scale was included, a familiar sounding-unfamiliar sounding scale, which did not enter into any of the composite scores. This scale was included to assess the relative familiarity of each of the stimulus items and always appeared at the bottom of the judgment forms.

All scales were translated into Japanese by a panel of five Japanese-English bilingual informants; subsequent to this first translation, a second panel of three bilinguals independently translated the scales from Japanese into English. The scales previously listed were all back-translated by the second set of informants into exactly the same form as given to the first panel of translators.

The seven-point scales were randomized as to order

TABLE 1

MATRIX OF CONSONANT AND VOWEL COMBINATIONS EMPLOYED IN THE EXPERIMENT

	p	n	č	š	g	
p		pin		piš		i
n	nip	nin				
č			čič		čig	
š			šič		šig	
g	gip			giš		
p	pep		peč			e
n		nen			neg	
č	čep			češ		
š		šen		šeš		
g			geč		geg	
p		pan			pag	a
n	nap			naš		
č			čač	čaš		
š		šan	šač			
g	gap				gag	
p				poš	pog	o
n			noč		nog	
č	čop	čon				
š			šoč	šoš		
g	gop	gon				
p		pun			pug	u
n			nuč	nuš		
č	čup				čug	
š	šup	šun				
g			guč	guš		

Note.—The 10 phonemic elements displayed in the table have the following orthographic equivalents: /p/ as in pat, /n/ as in nap, /č/ as in chap, /š/ as in shop, /g/ as in gap, /i/ as in cheap, /e/ as in gape, /a/ as in pot, /o/ as in note, /u/ as in shoot.

[3] Grateful acknowledgment is extended to Elizabeth T. Uldall of Edinburgh University for her expert assistance in the recording of the stimuli.

and position of the adjectival opposites. Booklets of 28 pages were assembled, each page duplicating the 16 scales, one page for each of 25 stimulus items plus three practice items used in the first of two testing sessions or three repeated items used in the second session.

Procedure

Instructions for the experiment were read to the subjects in the language appropriate to each language group. In the Japanese groups this was accomplished by a native Japanese experimenter, the author ostensibly present only to handle the tape recorder. The instructions essentially paralleled those suggested by Osgood, Suci, and Tannenbaum (1957, p. 82ff.).

Each subject served a total of 2 hours, each of the hours separated by at least a week's interval. Twenty-eight stimuli were presented at each session. The first three stimuli of the first session were practice items and were not analyzed. The first three stimuli of the second session were repeated items, chosen at random (the tape loops were shuffled in a large box) from the 25 experimental items presented in the previous testing. Subjects were tested in subgroups ranging in size from 5 to 20 members. The order of the 25 stimuli was randomized for each of the groups before presentation. The two sets of 25 items were obtained by dividing the 50 stimuli at random into two sets of 25 items each. This initial division of the items remained constant for both language groups, the randomization before each subgroup presentation within the two major language groups being used only on the two separate subsets of 25 items.

Each of the stimuli was repeated until all members of the subgroup had completed checking all 16 scales for that item. In those subgroups containing a few much slower subjects, a certain amount of speeded pacing was performed. No stimulus item, however, was repeated fewer than 15 times.

RESULTS AND DISCUSSION

Scale Intercorrelations and Factor Analyses

It will be recalled that the choice of particular scales to be grouped into composites for the purposes of later analysis was made quite independently of the data collected in this experiment. The present data were analyzed, however, to determine the adequacy of these a priori decisions. This analysis took the form of two sets of factor analyses and ancillary intercorrelations of the scale means for each of the language groups taken across all of the stimulus items. That is, each of the correlations is based on an N of 50 stimulus items for which the basic entry was the mean scale position of each of the subject groups as a whole. These intercorrelations were then factor analyzed using Thurstone's centroid procedure. A total of 12 factors was extracted from each of the language group scale intercorrelation matrices. Three of these factors were chosen for subsequent orthogonal rotation by means of the varimax criterion devised by Kaiser (1958). These factors were found to have approximate latent roots greater than 1.00 and were thus considered the major factors present in the correlation data. No other factors displayed approximate latent roots greater than one in either analysis.[4] The rotated factor structures for the two language groups are displayed in Tables 2 and 3.

It is strikingly clear that, despite some shuffled scales in the two analyses (the greater number occurring in the analyses of the Japanese data), these factors are easily identifiable as the well-established first three factors obtained from analyses of meaningful verbal material by Osgood et al. (1957). Although the choice of scales was deliberately dictated by the expectation that the same basic factors would appear in the present situation, the fact that they do maintain this structure when nonsense materials are being judged suggests that the same system of affective meaning can be extended to the new type of stimulus materials. In any case, these data justify the a priori decisions concerning the groupings of scales into composite scales. It remains to be seen whether some more extensive sampling of scales would have resulted in the factor structure identified here.

A comparison of the factor loading profiles for the two groups, using rank order correlations, was carried out to determine the degree of similarity of the two groups as to factor structure. Since the first two Japanese factors appeared to correspond most nearly to a reversal of those factors for the American analysis, these correlations were computed between Factor 1 of the American analysis, identified as the evaluative factor, with Factor 2 of the Japanese data, similarly identified; Factor 2 of the American data, the potency

[4] Raw factor analyses and analyses of variance tables have been deposited with the American Documentation Institute. Order Document No. 6767 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2
AMERICAN GROUP ROTATED SCALE CENTROID FACTOR LOADINGS

Scale	Factors		
	1	2	3
Good-bad	.82	−.08	−.04
Beautiful-ugly	.84	−.35	.11
Pleasant-unpleasant	.89	−.31	.16
Colorful-colorless	.78	.01	.45
High-low	.30	−.44	.75
Strong-weak	.12	.93	.16
Heavy-light	−.54	.68	−.40
Hard-soft	−.41	.54	.49
Powerful-powerless	.14	.90	.21
Thick-thin	−.37	.79	−.37
Large-small	−.13	.87	−.20
Passive-active	−.01	−.36	−.80
Quiet-noisy	.54	−.47	−.48
Slow-fast	−.23	.21	−.91
Cold-hot	−.60	−.09	.00
Familiar-unfamiliar	.70	.07	.16
Percentage variance	29.5	28.9	19.8

TABLE 3
JAPANESE GROUP ROTATED SCALE CENTROID FACTOR LOADINGS

Scale	Factors		
	1	2	3
Good-bad	.27	−.88	.22
Beautiful-ugly	.45	−.82	.26
Pleasant-unpleasant	.42	−.84	.18
Colorful-colorless	.84	−.26	−.32
High-low	.70	−.23	−.55
Strong-weak	−.46	.21	−.78
Heavy-light	−.90	.32	−.10
Hard-soft	.20	.46	−.72
Powerful-powerless	−.74	.24	−.52
Thick-thin	−.95	.14	−.14
Large-small	−.90	.05	−.27
Passive-active	.14	.14	.84
Quiet-noisy	−.11	−.47	.77
Slow-fast	−.87	.15	.33
Cold-hot	.81	.03	.21
Familiar-unfamiliar	−.04	−.80	−.12
Percentage variance	40.1	22.6	22.1

factor, was compared with its counterpart, Factor 1 of the Japanese; and the third or activity factors in each group were compared with one another. Loadings on the evaluative factors of the American and Japanese data correlated .89, the potency factor of the two groups correlated .74, and the activity factor in each group correlated .64. All correlations are significant at the .01 level. Thus it may be concluded, at least with respect to the profiles of factor loadings, that the two language groups use similar semantic dimensions.

Analyses of Variance

In view of the foregoing findings of reasonable independence of the three composite measures, the original decision to carry out independent significance tests on each of these composite measures appeared to be justified.

Assessment of the significance of the differences obtained in this experiment took the form of six separate four-variable classification analyses of variance, one for each of the three composite measures in each of the two language groups. The experimental classifications were those for consonant, vowel, position, and subject effects. The position variable represents the grand means for each of two equal subsets of the 50 stimulus items, each subset conforming to the experimental requirement of equal representation of all consonant elements before and after each of the vowel elements exercised for the stimulus list as a whole. Thus we have two miniature subsets of 25 stimulus items, each duplicating the experimental conditions imposed upon the original 50-item list. All variables with the exception of the subject effect were assumed to be fixed. The over-all model for the analysis is therefore of the mixed type. It was assumed that the interaction effects of the observations in which fixed and random effects enter are to be treated as random sampling variables. Tests of the random variable interactions and the subjects main effect were made utilizing the highest order interaction, or residual, in this analysis.

A summary of the significance levels found in all analyses is presented in Table 4. Only the .01 and .05 levels are reported, even though in most instances the F ratios were significant far beyond the .01 level.[5]

Inspection of the analyses for American subjects, displayed in Table 4, reveals that all main effects for all composite scores with the exception of the position effect are significant at the .01 level. The position effect reaches significance at the .05 level in the activity score analysis. Since all phoneme elements, as well as all consonant-vowel transitions, are

[5] See Footnote 4.

TABLE 4
SUMMARY OF SIGNIFICANCE LEVELS FOR ANALYSIS OF VARIANCE OF AMERICAN AND JAPANESE GROUP COMPOSITE SCORES

Source	American Group			Japanese Group		
	E	P	A	E	P	A
Consonant	.01	.01	.01	.01	.01	.01
Vowel	.01	.01	.01	.01	ns	.01
Position	ns	ns	.05	ns	ns	ns
Subject	.01	.01	.01	.01	.01	.01
V × C	.01	.01	.01	.01	.01	.01
V × P	.01	ns	.05	ns	ns	.05
V × S	.01	.01	.01	.01	.01	.01
C × P	.01	.01	ns	.01	.01	ns
C × S	.01	.01	.01	.01	.01	.01
P × S	ns	ns	ns	ns	ns	ns
V × C × P	.01	.01	.01	.01	.01	.01
V × C × S	ns	ns	ns	ns	.05	ns
V × P × S	ns	ns	ns	ns	ns	ns
C × P × S	ns	ns	ns	.01	.01	ns
V × C × P × S	—	—	—	—	—	—

represented an equal number of times in both sublists, the position effect provides a test of the adequacy of the experimental controls in randomizing the individual bias effects accruing to specific phoneme combinations. The activity composite score analysis indicates that this division was not adequate and represents a caveat that must be borne in mind in interpreting the effects of consonant position with regard to that score. The vowel × position and subject × position interactions provided a similar check on the adequacy of counterbalancing with respect to the vowel and subject determinations, respectively. The P × S interactions are all nonsignificant in these analyses. Two of the V × P interactions reached significance at or beyond the .05 level in the American data analysis. The bias reflected by these significant interactions is undoubtedly due to disproportionate effects of specific vowel-consonant combinations, the total effect of which is made more apparent because of the smaller number of items in the shortened sublists.

The higher order interactions were, as can be seen, uniformly nonsignificant in the American analyses, with the exception of the V × C × P effects. This latter interaction provided a test equivalent to a simple subjects by items analysis of variance and indicates that the stimulus items were reliably differentiated in the aggregate.

Inspection of the Japanese data analyses reveals that these results are in almost all regards parallel to those found for the American group. The most notable exception is the C × P × S interaction. The CVC pattern is relatively uncommon in Japanese. This fact could well account for the idiosyncratic subject judgments of the consonant-position combinations and hence for the significant variance contribution of this interaction. Notice should also be taken of the nonsignificant vowel effect for the potency composite measure in this group.

With the specific exceptions apparent in the table, the two sets of analyses indicate reliable differentiation of the phoneme elements, consonant positions and unique stimulus items. In addition, even though subjects were reliably differentiated with respect to judgments of these phoneme elements, such differences were not great enough to obscure the obtained differences between the elements themselves. Further, the absence of significant highest order interactions and stimulus-list subset biases, with several notable exceptions, lends credence to the utility of the experimental design.

Familiarity Judgments

In order to assess the extent of the influence of the subjects' judgments of the familiarity of the stimulus items upon the composite measure judgments, the familiarity scale scores were correlated with each of the composites. These correlations took the form of Pearson product-moment coefficients, for which the N was the 50 stimuli. The correlations for the American data between familiarity judgments and the evaluative, potency, and activity composites were: .59, −.15, and −.12, respectively. In the Japanese data, correlation coefficients between familiarity judgments and evaluative, potency, and activity composites were: .50, −.15, and −.02, respectively. In addition, the mean familiarity judgments indicated that the stimuli were in truth relatively unfamiliar, although significant internal variation in this regard was present. If the subjects' judgments of these nonsense stimuli had been conditioned by some generalization from close real-word associates, the covariation of the familiarity and composite measures should have been much more extensive. The data, however, support a conclusion that

attributes the meanings assigned to these stimuli on the basis of inherent symbolic content other than, or at least beyond, that which might accrue to generalization from real words more familiar to the subjects.

Composite Scale Means

The preceding results indicate that we are justified in examining the stimulus configurations and phoneme element means in some detail in order to determine the precise nature of the symbolic content obtained. The individual phoneme element trends are displayed in Figures 1, 2, and 3. These displays are ordered so that the stimulus elements follow the organogenetic descriptions of the elements with regard to articulatory position from "front" to "back," in the classical sense of these terms.

Inspection of Figures 1 and 2 indicates, in general, that the more anterior consonant or vowel elements are judged to be more positive in evaluation and more negative in potency than the more posterior elements. The functions are hardly monotonic in character, although the direction of the change as a function of the stimulus dimension is surprisingly regular in character for the evaluative and potency composites. The activity composite means are much more erratic in terms of the dimension employed. This composite measure, on the other hand, accounted for the least amount of total judgment variance of the three composites as determined by means of the factor analyses.

Of the alternative orderings of the phoneme elements in terms of physical correlates, the organogenetic basis is probably not the most desirable. Instead, since the acoustic endproduct of the articulatory modifications corresponds more closely to the proximal stimulus giving rise to the perception of these stimuli, it might seem reasonable to turn to this type of description as a basis for the stimulus ordering. Fairbanks (1960) has dramatically demonstrated that the organogenetic description, however, is almost perfectly correlated with the acoustic description in terms of formant locations. His vowel plots using Formant 1 and 2 coordinates indicate that the two descriptive systems are highly redundant. Although the stimuli were not actually submitted to any harmonic analysis, it is possible to use data collected by Fairbanks and others

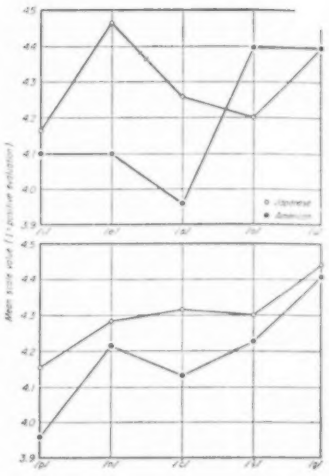


FIG. 1. American and Japanese group mean evaluative composite scores for vowels (upper) and consonants (lower).

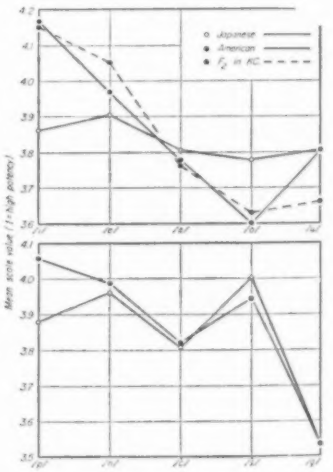


FIG. 2. American and Japanese group mean potency composite scores for vowels (upper) and consonants (lower).

in assessing the degree of covariation of these judgments with other acoustical analyses of the phonemes selected. The upper display of Figure 2 shows the relative values of Formant 2, after rescaling, as taken from data collected by Holbrook (1958). The most striking difference between the two descriptive systems in terms of the judgmental data collected in

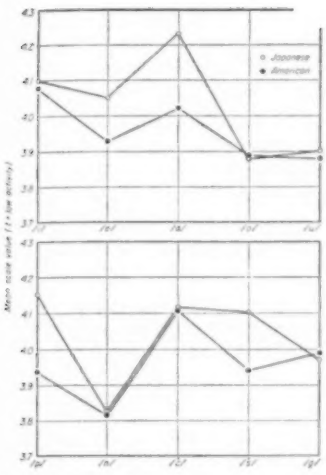


FIG. 3. American and Japanese group mean activity composite scores for vowels (upper) and consonants (lower).

this experiment involves the return of the potency judgments for the vowel /u/ to a higher level, ("weaker" judgments) giving the function a J-shaped character. The F_2 formant values follow this function at all points for the American data. (It will be recalled that the Japanese data did not disclose a significant vowel effect for this composite measure.)

At least one reasonable interpretation for the obtained effect on this composite measure derives from the nearly universal correlation between high pitches and small objects to be found in nature. Thus by metaphorical extension, to a correspondence between small objects and weakness. A previous experiment reported by Solomon (1959), in which passive sonar signals were evaluated on a set of semantic differential scales similar to those used in this study, obtained almost identical results. Scales that were classified as contributing to a potency composite were found to distinguish between sonar signals having a preponderance of energy in the lower octave bands from those with energy concentrated in the higher bands. The direction of the correspondence was as reported in this study, i.e., the greater the low frequency energy in the signal, the "heavier" and "stronger" the judgments, and vice versa.

Cross-Linguistic Correspondences

In order to assess the over-all correspondence of the two language groups in phonetic symbolism tendencies, product-moment correlations were computed for each of the composite scores utilizing all determinations available within each group, i.e., the individual mean judgments for each of the stimulus items. These correlations were found to be .46, .53, and .54 for the evaluative, potency, and activity composites, respectively. All correlations are significant at the .01 level with 49 *df*. The summary measures for each of the composite scores for each of the five vowel and consonant phoneme elements, as displayed in the foregoing exhibits, were also compared across the two language groups. The evaluative composite correlation coefficient for the vowel elements was computed to be .23 and that for consonants .95. The coefficient computed for the vowel elements is not significant while the consonant correlation reaches significance at the .01 level. The comparable correlations for the potency composite were found to be .57 and .91 for the vowel and consonant elements, respectively. The activity composite correlations were .61 for the vowels and .64 for the consonants. All correlations for these second two composites are significant at the .05 level or beyond. As can be seen, the coefficients for the consonant correlations are uniformly higher than those for the vowels. If we assume for a moment that phoneme element judgments may be conditioned in part by some imaged articulatory posture characteristic of the sound elements, the foregoing result suggests an interesting conclusion: Since the consonants are more easily discriminated with respect to articulatory production, i.e., they are characterized by more easily identifiable muscle features of more diverse kinds, it is possible that the higher degree of correspondence obtained for these elements reflects the presence of this greater number of discriminable cues.

In conclusion, it seems clear that, although this study does not afford direct evidence regarding the existence of phonetic symbolism in natural languages—evidence difficult to obtain in any study—the magnitude of the differences in meanings assigned to these stimuli and the similarity of these meanings across the two language groups provides at

least indirect evidence of its presence and generality.

SUMMARY

An experiment is reported in which a set of 50 nonsense stimulus items of CVC shape were judged on a set of 16 semantic differential scales by two groups of subjects with contrasting linguistic backgrounds. The subject languages used were Japanese and English. Experimental stimulus controls took the form of an incomplete latin square design in which each of five consonant elements preceded and followed each of five vowel elements an equal number of times with all elements represented and equally frequent in the final matrix.

The mean judgments for all subjects in both groups indicated a general trend toward evaluating the front vowels and consonants as "pleasant" and "weak" and the back vowels and consonants as more "unpleasant" and "strong."

Four-way analyses of variance indicated that, in the main, the phoneme elements, the consonant positions, and the phoneme combinations were reliably differentiated in the aggregate by the subjects of both groups.

Data collected on the familiarity of the stimulus items showed significant internal variation but was shown to covary with evaluative scores only.

Centroid factor analysis of the scale intercorrelations revealed the presence of three factors that accounted for a high percentage of the total judgment variance for both groups. These factors were easily identifiable as corresponding to previously observed dimensions and were found to have a high degree of similarity across the two language groups.

It was concluded that the materials had expressive symbolic value accruing to their inherent phonetic content and not to any meanings via real-word associates. These affective meanings were found to bear consistent, lawful relations to the phonetic properties of the sounds. The fact that these meaningful differentiations and their relations to phonetic properties proved to be highly similar across two contrasting linguistic groups suggests that the laws governing phonetic symbolism may have a universal character.

REFERENCES

BRACKBILL, YVONNE, & LITTLE, K. Factors determining the guessing of meanings of foreign words. *J. abnorm. soc. Psychol.*, 1957, **54**, 312–318.

BROWN, R., BLACK, A., & HOROWITZ, A. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, **50**, 388–393.

FAIRBANKS, G. *Voice and articulation drillbook*. New York: Harper, 1960.

HOLBROOK, A. An exploratory study of diphthong formants. Unpublished doctoral dissertation, University of Illinois, 1958.

KAISER, H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

MALTZMAN, I., MORRISETT, L., JR., & BROOKS, L. O. An investigation of phonetic symbolism. *J. abnorm. soc. Psychol.*, 1956, **53**, 249–251.

OSGOOD, C. E., SUCI, G., & TANNENBAUM, P. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

SOLOMON, L. N. Search for physical correlates to psychological dimensions of sounds. *J. Acoust. Soc. Amer.*, 1959, **31**, 492–497.

(Received June 6, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 631-639

PANCULTURAL FACTOR ANALYSIS OF REPORTED SOCIALIZATION PRACTICES[1]

LEIGH MINTURN TRIANDIS[2] AND WILLIAM W. LAMBERT
University of Illinois *Cornell University*

Most factor analyses of child training practices have been done on data obtained from parents in one culture, usually USA or Europe (e.g., Takala & Takala, 1960). This paper reports a factor analysis of the stated socialization practices of mothers in six different cultural settings obtained through systematic interviews. The analysis was done in this form for the purpose of reducing a large number of scales in order to test antecedent-consequent hypotheses, but the results have value in their own right.

The identification of pancultural factors may suggest ways in which studies of child training practices may be simplified by focusing the researcher's attention on major sources of differences in mothers' statements regarding their practices (and possibly differences in the actual practices themselves). To the extent that the factors extracted are clear and are determined by differences between mothers rather than cultural differences alone, the factors can be used as dimensions in the direct comparison of practices in different societies or cultural contexts. Finally, the factors extracted may assist in formulating more relevant theories concerning the functional prerequisites for any society, or, more modestly, concerning sets of antecedent or consequent variables of personality or social structure.

[1] This paper is a report on a portion of work done under grants from the Social Science Research Council, the Ford Foundation (Behavioral Science Division), and NIMH (Project M2331) to the Cornell-Harvard-Yale Study of Socialization in Five Cultures, directed by John W. M. Whiting, Irvin Child, and William W. Lambert. Other aspects of work reported to date can be found in Whiting, Child, Lambert, Fisher, Fisher, Minturn, Maretski, Maretski, Nydegger, Nydegger, Romney, and Romney (1955), Lambert (1960), and Whiting and Whiting (1960).

[2] The authors wish to express their gratitude to Jack and Ann Fisher, Thomas and Hatsumi Maretski, William and Corinne Nydegger, Kimbal and Romaine Romney, and Robert and Barbara Levine for the use of their data. These investigators gathered the data in the Northeastern United States, Okinawa, the Philippines, Mexico, and Africa, respectively. The Indian data were gathered by Leigh Minturn Triandis.

The problems involved here are many, and not all of them will be dealt with in the present paper. There are issues of the degree to which what mothers report as their practices are in fact their practices; there are issues concerning the relationship of intracultural factor structures to the pancultural factors reported here; there are problems arising from translation and back translation of the questionnaire and responses, etc. Many of these issues are to be dealt with in subsequent reports; some are not resolvable in the present state of our methods. Our present aim is to report what was done and what emerged from the analysis, along with a preliminary evaluation of results.

Field Procedure

Field teams (usually a husband and wife) were trained and sent out to relatively homogeneous communities in North India, Okinawa, Mexico (Indian), the Philippines, the Northeastern United States, and to an African tribe in Kenya, in order to conduct intensive studies in techniques of socialization and the behavior of children. An attempt was made to use several different methods in each cultural setting, but also to study the handling of several behavior systems that could be viewed as common to all the societies. One of the instruments used was a standard interview (with probes) which was prepared in the United States, and then translated into the native language. In some cases the questions were altered in order to make the questions fit the concepts and practices of the several cultural settings. The answers of the mothers (or mother-surrogates) of a carefully sampled group of children were recorded *verbatim* to the 21 classes of questions, with an average of four subquestions per class. The total number of mothers interviewed was 133.

The samples of mothers are reflected in Table 1, in terms of the distribution of age and sex of the child that each mother talked about. The ages ranged from 3 to 10 years, each child came from a different family, and

an attempt was made to counterbalance birth order and to ensure that each sample child was fairly normal physically.

All of the interviews were translated into English with the aid of bilingual informants where necessary. They were all sent to Cornell and analyzed by trained coders and raters.

ANALYSIS PROCEDURE AND RELIABILITY

Since it was impossible to use all of the information obtained from the mother interviews in the factor analysis, scales were selected on the following bases: (*a*) a minimum of "no answers" in all six subsamples; (*b*) relatively high (estimated) coder reliability; (*c*) theoretical relevance, including relevance to past factor analyses. Not all of these considerations could be maximized, of course. The scales finally retained for the factor analysis are listed in Table 2, which tap a number of aspects of socialization practices: the handling by socializing agents (particularly the mother) of a number of different behavior systems, aspects of caretaking, forms of discipline, some kinds of rules, and the apparent or reported communication of these rules, and the sex and age of the child.

Three types of scales are included in Table 2. Scales 1 through 3 represent a placement of the mother's verbatim answer into a category implied by the following anchor points of a nine-point scale: (1) Mother (or surrogate) physically punishes child or deprives child of privileges; (3) Mother reprimands or teaches; (5) Mother withholds approval or ignores; (7) Mother gives approval or gives emotional support; (9) Mother gives reward or gives active help. The answers of the mothers were placed on intermediate points of this scale when necessary, and if the mother's answer indicated that her response encompassed items on two points of the scale (e.g., "Sometimes I scold him; sometimes I just ignore it") the answer would be an average of the two: the example would be coded 4 on the nine-point scale.

The second type of scale in Table 2 involves the seven-point scales numbered 4 through 18, which the coders rated on the basis of reading the entire verbatim protocol of the mother's interview. Point 7 represents the high rating and Point 1 the low rating.

TABLE 1
THE NUMBER OF CHILDREN

Country	Girls 3–6 yrs.	Girls 7–10 yrs.	Boys 3–6 yrs.	Boys 7–10 yrs.
USA	6	6	6	6
India	6	6	6	6
Okinawa	6	6	6	6
Philippines	6	6	5	6
Mexico	6	5	6	5
Africa	4	4	4	4

TABLE 2
SCALES USED IN THE FACTOR ANALYSIS

1. Degree to which mother, when busy, responds to child's requests for help
2. Degree to which mother is positive when child fights with other children
3. Degree to which mother is positive or nonpunishing when child becomes angry while being scolded
4. Proportion of caretaking done by mother
5. General warmth of mother
6. Mood variation of mother's warmth
7. General hostility of mother
8. Mood variation of hostility
9. Degree that mother's warmth is contingent upon child's actions
10. Degree that mother's hostility is contingent upon child's actions
11. Amount of communication of rules
12. Consistency of aggression rules
13. Amount of praise
14. Consistency of mother's follow-through on nonroutine demands for obedience
15. Frequency and intensity of physical punishment
16. Degree to which privileges and gifts are contingent upon child's behavior
17. Reward for retaliatory aggression to peers
18. Degree to which mother is aggressive when child is angry or aggressive
19. Amount of time that mother cares for child
20. Amount of time that the mother cared for child when child was a baby
21. Amount of time adult other than the mother cared for child when child was a baby
22. Age at which child began to dress self
23. Age at which child began to play away from the house
24. Total frequency of chores
25. Total number of chores
26. Degree to which mother expects immediate obedience
27. Sex of child
28. Age of child

The third class of scales are those numbered 19 through 28, which includes a variety of scales (age, time, frequency, etc.) used to place the answers of the mothers to individual questions. The nature of the scale used in each case is fairly obvious from the name given it in Table 2.

Not all mothers answered all questions in

such a way that their answers could be placed on all of the scales, and this raised problems for the computing machine, which cannot conveniently handle such omissions. The problem was overcome by entering the average response given by all the other mothers in the given culture in the place of the "no answer."

The reliability of the codings and ratings is reported in Table 3. A fairly random sample of protocols (from different societies) was coded by two coders, the number of the sample involved being given in the second column of Table 3 opposite the relevant scale number. The next column reports the number of times only one coder entered a judgment (a form of error). The next to last column contains the percentages of disagreement exceeding one point on the relevant scale. The final column contains the Pearson correlation coefficients between the two raters' judgments. These coefficients reflect the errors due both to the fact that one rater made a judgment and the other did not, and to disagreements in judgments. The first source of error is quite important, since the reliability of Scale 23 rises to 1.00 when only cases where both raters made a judgment are used. The coefficients are not corrected for attenuation. On the basis of these considerations the coder reliability was regarded as adequate for the task.

The factor analysis was conducted with the Illiac computer, with the data from all six cultures put together to produce a correlation matrix of the variables listed in Table 2. The matrix was then analyzed, using a centroid analysis with iterated communalities. The factor scores were then rotated for orthogonal simple structure with the Illiac varimax rotation program. Ten factors were extracted, seven of which appear to present few problems of interpretation.

TABLE 3
DATA REGARDING SCALE RELIABILITY

Scale Number	Total number of entries	Number of single answers	% Disagreement greater than one point	Correlation coefficient between two raters
1	19	1	0	.77
2	6	2	50	.53
3	9	0	1	.68
4	17	0	24	.49
5	17	1	19	.55
6	16	3	14	.59
7	16	4	13	.74
8	17	5	17	.67
9	17	2	30	.49
10	16	1	18	.47
11	17	6	3	.81
12	17	1	6	.54
13	17	1	19	.45
14	17	2	13	.67
15	17	0	8	.49
16	17	0	13	.53
17	17	0	5	.51
18	17	4	7	.61
19	23	0	26	.73
20	23	0	39	.71
21	23	0	35	.85
22	22	0	0	1.00
23	21	2	0	.29[a]
24	23	0	0	.96
25	23	0	0	.98
26	23	0	56	.65
27	23	0	0	1.00
28	23	0	0	1.00

[a] If computed using double entries only, 1.00.

RESULTS AND INTERPRETATION

The reference vector factors that resulted from the varimax rotation are presented in Table 4. For the purposes of factor definition, we will use an arbitrary cutting point of .40, but will discuss the relation of scales with lower loadings to the factor definition, particularly loadings from .30 to .40, since such loadings may help to clarify the factor.

Table 5 presents the amount of both the extracted variance and of the total variance accounted for by each of the 10 factors. Column 1 shows the amount of variance accounted for by all 10 factors that is attributable to each particular factor. The second column shows how much of the total variance of the original correlation matrix is accounted for by each factor. All 10 of the factors extracted account for 48% of the total matrix variance. Although we would expect that a good deal of the remaining variance is due to error of one form or another, some of it may well be involved in factors unique to the cultural settings studied. This problem is to be discussed in later reports.

Before discussing the factors of Table 5 in more detail some readers will wish to peruse Table 6 which presents the results of an analysis of variance of the degree to which the factor variance (in the first seven of the factors) is due to variance within societies (that is, to differences in mothers or children) as compared to variance between societies (due, thence, to differences in the cultural means).

TABLE 4
VARIMAX ROTATION ON TEN CENTROID FACTORS USING ITERATED COMMUNALITIES
(Overall Analysis)

Scale Number	Factor Number									
	1	2	3	4	5	6	7	8	9	10
1	+.14	+.06	−.06	+.04	−.09	−.02	+.09	−.08	−.08	−.52
2	+.11	+.00	−.05	+.07	+.06	+.03	+.14	+.02	+.48	+.15
3	−.15	−.16	+.08	+.02	−.06	−.76	−.17	+.05	+.11	−.27
4	+.06	+.03	−.02	+.04	+.74	−.02	+.06	+.03	+.07	+.11
5	−.18	+.61	−.09	−.07	+.03	+.09	+.04	−.02	+.32	−.05
6	−.02	−.05	+.02	−.04	+.27	+.01	+.74	−.10	+.08	−.17
7	−.11	−.69	−.09	−.02	−.01	+.05	+.34	−.02	+.14	+.10
8	−.08	−.20	−.07	+.03	+.10	+.10	+.91	−.01	+.01	−.02
9	−.18	−.21	+.25	−.03	−.14	+.17	+.26	−.50	+.12	+.19
10	+.07	−.08	+.16	−.22	−.29	−.09	−.21	−.29	+.07	+.08
11	+.04	+.03	+.24	−.29	+.21	+.22	−.05	+.03	+.28	−.02
12	+.04	+.02	+.70	−.21	−.08	−.17	−.09	−.09	−.07	+.04
13	−.11	+.52	−.03	−.31	−.10	+.09	+.05	−.13	+.15	+.12
14	−.05	+.00	+.49	+.09	+.03	+.36	+.00	−.04	+.23	−.09
15	−.04	−.58	+.18	+.04	−.11	+.33	+.19	−.29	+.05	+.10
16	+.11	−.08	+.14	−.09	−.22	+.39	−.11	−.17	+.05	+.28
17	−.10	+.03	−.68	+.16	+.24	−.03	−.01	+.03	+.16	−.09
18	+.03	−.28	+.11	−.28	+.05	+.64	+.09	−.02	−.14	−.20
19	−.02	−.02	−.09	+.12	+.54	+.04	+.15	−.10	+.01	−.02
20	+.01	−.07	−.03	+.72	+.08	−.02	−.10	+.03	+.07	−.12
21	+.01	+.05	+.03	−.46	−.08	+.04	−.06	+.02	+.10	−.05
22	+.03	−.04	+.11	−.02	−.13	−.01	−.02	+.53	−.00	+.14
23	+.29	+.02	+.01	+.15	+.04	+.36	+.00	+.09	+.23	−.12
24	+.64	−.05	+.02	+.14	+.05	+.24	−.02	−.22	−.04	−.12
25	+.69	−.05	−.01	−.16	+.00	+.08	−.09	+.06	+.01	+.06
26	−.02	−.03	+.37	+.11	+.03	+.06	−.01	+.07	+.02	+.01
27	+.01	+.06	−.01	−.10	.00	−.04	−.03	−.03	+.35	−.01
28	+.56	+.05	+.12	+.10	+.01	−.12	+.04	+.22	+.21	−.17

The variance is preponderantly determined by within-society variance in all cases. In only the case of Factor 6 does the between-society variance approach the within-society variance. This is evidence that we are dealing with factors that have pancultural meaning and may be useful in more than one societal context for the study of socialization processes.

Factor 1: Responsibility demands on the child. This first factor is easily interpreted, since it has to do with responsibility demands on the child. The ratings include not only the chores assigned to the child by the mother, but also those assigned by others. Therefore this factor deals with the general pressures on children for doing things to help the family get its work done, and some students of groups would not be surprised that this task orientation should emerge at once from a factor analysis of socialization processes. The factor may have relevance, then, to group processes as well as to personality development, as is true of most of the factors below.

We can characterize the children who have high scores on this factor as follows: they do a relatively large number of different chores (Scale 25, +.694); they do chores relatively more frequently (Scale 24, +.64) and, as one might expect, they are to some extent relatively older (Scale 28, +.56).

These are the only scales with loadings of greater than .30, but it is interesting that Scale 23, which relates to the age at which children begin to play away from home, has a loading of +.29 in Factor 1. This may be because the chores take a child away from home early, thus giving him (to some extent) greater freedom in play as well.

Factor 2: Use of positive affect by mother. Factor 2 is also easy to interpret, having to do with what is often referred to as "maternal warmth." We have chosen to label this factor "use of positive affect by mother," however, since instrumental as well as expressive aspects seem to be involved. Mothers at the warm end of this dimension are characterized by low general hostility (Scale 7, −.69), high general warmth (Scale 5, +.61), relatively high use of praise (Scale 13, +.52), and relatively low frequency and intensity of physical

TABLE 5

PERCENTAGE OF EXTRACTED AND OF TOTAL VARIANCE REFERABLE TO EACH OF THE DEFINED FACTORS

Factor	Percentage of the extracted variance	Percentage of the total variance
1	11.1	5.2
2	12.7	6.0
3	12.1	5.7
4	9.5	4.5
5	9.4	4.5
6	13.1	6.2
7	13.3	6.3
8	6.9	3.2
9	6.3	3.0
10	5.4	2.6
Total	100.0	48.2

TABLE 6

DETERMINANTS OF VARIANCE OF FACTOR SCORES

Factor	Percentage of within society variance	Percentage of between society variance
1	94.5	5.5
2	83.7	16.3
3	67.2	32.8
4	92.8	7.2
5	73.5	26.5
6	52.9	47.1
7	76.3	23.7

punishment directed toward the child (Scale 15, −.58).

These scales are consistent in meaning that mothers with high loadings are warm rather than hostile with their children, and that they control them through praise rather than through punishment and threat of punishment. The only other scale that approaches an interesting level of loading is Scale 18, which has to do with the degree to which the mother is aggressive when the child is angry or aggressive (−.28), which is certainly consistent in tone.

Factor 3: Control of the child's peer aggression and obedience. This factor clearly deals with an important aspect of aggression control, focusing on the peer relationship, but also dealing with expectations and family rules regarding obedience. Mothers who have high scores on this factor may be characterized as having consistent rules regarding aggression (Scale 12, +.70), they tend to punish the child's retaliatory aggression directed toward the instigating peers, and the mother tends to be relatively consistent in following up on her nonroutine demands for obedience (Scale 14, +.49). It is consistent that these mothers also tend to expect *immediate* obedience (Scale 26, +.37) from their child.

There is no inconsistency in finding two low positive loadings in Scale 11 (+.24), reflecting the amount of communication of rules in general by the mother (certainly aggression rules are a component here, though the persistent articulateness of the mother may not be high), and in Scale 9 (+.25), contingency of mother's warmth. A mother who had rules in the areas of aggression and obedience and followed up on them would be expected to make her warmth at least somewhat dependent upon the child's obedience and his nonretaliatory behavior to peers.

In defining this factor, in which the loading on reward for retaliation (toward peers) looms so large, it is interesting that Scale 2, "Degree to which mother is positive when child fights with other children," has essentially a zero loading on this factor, but defines (if anything does) Factor 9. This may suggest a distinction between retaliating when one has been picked on as compared to starting a fight (or, more vaguely, just "being in a fight"). One might expect these to differ, but the surprise arises in their orthogonality. It is also interesting that the mother's reaction to her child's retaliation to peers is not correlated with her own reaction to the child's aggressiveness while being scolded (Scale 3), nor to the degree to which she herself uses aggression as a reaction to the child's aggressiveness (Scale 18). These last two scales define Factor 6, and are therefore orthogonal to Factor 3. Panculturally viewed, then, aggression toward peers is a different thing, and is socialized differently, from aggression toward parents or other socializing agents.

Factor 4: Extent of mother's caretaking of the baby. This factor is clearly involved with caretaking. Mothers who score high on this factor can be characterized by the fact that the mother had a higher proportion of the responsibility for her child when he (or she) was a baby (Scale 20, +.72) and that other adults were relatively responsible for at this activity (Scale 21, −.46).

There are a number of other scales that have low, but suggestive, loading on this factor. There seems to be some tendency for mothers

high on Factor 4 to show "flat" affect to their children, since they are less aggressive when the child is angry or aggressive toward them (Scale 18, −.28), and they use praise less (Scale 13, −.31). It is also possible that these mothers, as a result of long contact with their *babies*, can assume more about their *child's* behavior, since they do less communicating of rules (Scale 11, −.29), they have less consistent aggressive rules (Scale 12, −.21), and their own hostility is somewhat less contingent upon the behavior of the child (Scale 10, −.22). Perhaps all these low loadings can be interpreted as displaying the fact that if the mother has had long association with her baby, she needs to give less overt information to the baby when it has become a child. The interesting thing here, then, is that this holds for affectional as well as instrumental information.

Note that the scales which reflect the proportion of the caretaking done by the mother (Scale 19 in Factor 5, as well as Scales 20 and 21 in Factor 4) reflect answers to the questions, "Who usually cares for your child?" and "Who did this when he was a baby?" The answers were scaled from "never" to "always." These scales do not reflect, or only reflect imperfectly, the total *amount* of time the mother spends with the child. They do reflect the degree of responsibility for care of the child. Otherwise we would expect higher correlations with scales that reflect the personality of the mother more directly.

Factor 5: Extent of mother's caretaking of the child. This factor clearly represents the proportion of the caretaking of the *child* that is done by the mother at the time of interviewing (Scale 4, +.74; Scale 19, +.54). It is interesting that this matter is orthogonal to Factor 4, which deals with the caretaking of the baby.

There are only two low loadings with this factor, which suggest ways in which caretaking has harassing effects upon at least some mothers: the heavy caretakers tend to have more variation in warmth due to mood (Scale 6, +.27) and to have hostility that is not contingent upon the child's performance (Scale 10, −.29). Perhaps much care of a child gives mother license for self-expression (though the evidence here is meagre).

Factor 6: Control of mother directed aggression and disobedience. This factor has to do with the control of negative feelings in the child that are directed toward the parent. The defining scales suggest that mothers who have high scores on this factor tend *not* to be positive or nonpunitive when the child becomes angry when scolded (Scale 3, −.76), and that they tend to be more aggressive when the child is angry or aggressive toward the mother (Scale 18, +.64).

There are four scales with low loadings (in the .30 to .40 range) which help to define the factor. These loadings suggest that mothers who are high on this factor tend to make consistent demands for nonroutine obedience (Scale 14, +.36), to use physical punishment more frequently and intensely (Scale 15, +.33), to control the child in part through making gifts and privileges contingent upon his good behavior (Scale 16, +.39), and (for some reason) to permit play away from home at a relatively later age (Scale 23, +.36).

The emergence of this factor certainly suggests the existence of a number of important dimensions of aggression that must be partialed out in cross-cultural analyses. If such is in fact the case, and to the degree that one views personality development from the perspective of learning theory, then these distinctions suggest that later peer aggression may be somewhat orthogonal to aggression directed toward those in authority.

Factor 7: Emotional instability of the mother. This factor is strongly defined by Scale 8: "mood variation of hostility" (+.91), but there is also a strong loading in the Scale 6 of +.74, reflecting the mood variation of the mother's warmth as well. The lower loading of +.34 on Scale 7, "general hostility of the mother" somewhat broadens the meaning of the factor. This kind of factor may be very important in socialization, since a mother high in it may provide a difficult "target" for identification and an aggressive role model, and the study of such persons may provide clues to familial and even cultural sources of stress.

The remaining three factors are less well defined than are the seven just discussed. Since the loadings are relatively low, and they account for rather little of the variance, the factors should be viewed as very tentative indeed.

Factor 8. This factor is defined by two

scales, which suggest that mothers who score high tend to give warmth to the child which is not contingent upon his good performance (Scale 9, −.50), and that the children of these mothers are able to dress themselves at a relatively later age (Scale 22, +.53). We may thus be dealing with a weak factor suggestive of "mothering" or of "noncontingent nurturance." The weak loadings are consistent with such an interpretation, suggesting that the mother keeps the child from responsibility (Scale 24, −.22, dealing with total frequency of chores) and that she does not try to use punishment to control behavior (Scale 10, −.29, suggesting low contingency of mother's hostility; and Scale 15, −.29, showing somewhat less tendency to use physical punishment in socialization).

Factor 9. This relatively weak factor may be an artifact, since, though it deals with parental attitudes toward peer aggression it is not loaded on Scale 17, "reward for retaliation" directed toward peers. There may, however, be some sense to this factor if it is viewed as "peer aggression permissiveness" (Scale 2, +.48, reflecting the "degree to which mother is positive when child fights other children"), tied somewhat to boys (Scale 27, +.35). It is interesting that the only other loading above .30 is the "general warmth of the mother" (Scale 5, +.32), which gives a picture of an expressive, positive mother, who permits her boy to be expressive as well. This suggestion can be checked from other data to be reported elsewhere.

Factor 10. This weak factor has only one defining scale, and should probably be ignored (Scale 1, −.52, reflecting the fact that this mother does not give help to the child when she is busy). As a hunch, however, we may be dealing with a weakly adumbrated factor of "mother irritability," since there is a weak loading on Scale 3 (−.27) suggesting that the mother tends also to be punitive when the child is angry when scolded, and on Scale 16 (+.28), showing that she gives gifts and privileges contingent on the child's good behavior.

DISCUSSION

We have already suggested possible contexts in which the list of factors we have presented may prove useful. We have also suggested some of the additional problems regarding these data which await further analysis.

For a number of reasons, it is difficult to relate our pancultural findings to previous research on parent behavior. First, all of the analyses available are intracultural rather than cross-cultural. Second, researchers in this area have not been uniform in their choice of the types of scales. Resemblance of factors, then, sometimes rests on one or two scales in one factor, which seem similar in meaning to scales in another. Third, there is diversity in the types of analysis. Baldwin, Kalhorn, and Breese (1949) for instance, analyzed their results by means of cluster analysis; Sewell, Mussen, and Harris (1955) used the multiple group method in extracting their factors and extracted correlated factors by this method; Schaefer and Bell (1957) used the centroid factor analytic procedure, but oblique rotation; Zuckerman, Ribback, Monashkin, and Norton (1958) describe unrotated factors and compare their results to unrotated factors obtained by Schaefer and Bell (1958); Milton (1958) used a procedure most comparable to ours, that is, centroid analyses and orthogonal rotation. Finally, comparison of factors obtained on different samples must be to some extent subjective, despite the mammoth endeavors of Ahmavaara (1957) and others, since there is as yet no definitively precise method of comparing factoral results obtained on different samples.

However, since there seems to be some similarity of factors across various American samples, it seems reasonable, despite the difficulties involved, to consider which of these factors have some pancultural basis.

The factors that have commonly emerged in studies of American samples have been versions of warmth factors and permissive-restrictive factors. It is not surprising, then, that our warmth factor (Factor 2) and restriction factors (Factor 3, Factor 6) are the ones which have the most comparability with the results of other studies.

Factor 2 (Use of positive affect by the mother) in the present study has a positive loading on maternal warmth and a negative loading on maternal hostility. Mothers high on this factor also praise their children and

tend not to use physical punishment. Milton (1958) also distinguishes a warmth factor (Factor C) which includes, among other scales, affectionate demonstrativeness and praise for table manners. Like our warmth factor, Milton's is orthogonal to his permissiveness factor. Zuckerman et al. (1958) and Schaefer and Bell (1958) find a hostile rejection factor in their unrotated area, which might be considered the opposite end of a warmth factor. Baldwin et al. (1949) also find a warmth factor in their cluster analysis.

In our study we have isolated two (possibly three) factors concerned with aggression training. One (Factor 3) is concerned with restrictions on aggression to peers and obedience training, the other (Factor 6) is concerned with restrictions on aggression directed at the mother. Milton (1958) finds one aggression factor (Factor E) which includes treatment of aggression directed at peers and at parents. He finds that the factor loading for permissiveness toward peer aggression is correlated with punishment of aggression to parents. We find them to be independent, appearing, as they do, on separate factors. The comparison with Milton's data is further complicated by the fact that he also finds a permissiveness factor (Factor A) broader in scope than aggression training and obedience, but including these scales. Thus in Milton's analysis, the scales measuring restrictions on aggressive behavior appear in two factors, but always together, whereas in our data they appear in two separate factors. Schaefer and Bell (1957) find a factor which they call Over-Possessiveness (Factor IV) and Zuckerman et al. (1958) find one called Authoritarian control (Factor T), both of which include a scale on restrictions on aggression. Baldwin et al. (1949) find a general restrictiveness cluster and Sewell et al. (1955) find a rather diverse factor, which they leave unnamed (Factor 7) but includes strict obedience training. While these findings have only tangential resemblance to our aggression training factors, they do seem to be generally concerned with dimensions of strictness of discipline.

Our emotional instability factor (Factor 7) also has some parallels in other studies. Schaefer and Bell (1957) find a factor called Hostile Rejection of the Homemaking Role, and Zuckerman et al. (1958) find a Hostility-Rejection factor, both of which include a scale of irritability. Baldwin et al. (1949) also find an Emotionality cluster. Milton's (1958) General Family Adjustment Factor, which includes hostility of the father and mother's child rearing anxiety, may represent a related pattern.

It remains to mention that Sewell et al. (1955) find a factor (Factor 4) which they call Non-Punitive Treatment that includes ignoring the child's neglect of jobs. This might be similar to our first factor, which is concerned with responsibility training.

Finally, none of the American studies we have reviewed include scales on the proportion of caretaking that is done by the mother. Therefore we find no factors comparable to our Factors 4 and 5 that deal with this area. These scales differ from most scales in our study and the other studies here reviewed in that they do not purely represent either attitudes towards child rearing practices, or personality variables. However, such factors may certainly be related both to attitudes of the parents and to behavior of the child. The variable of time spent with the child has been extensively investigated with institutional children and the children of working mothers. It seems rather odd, in view of the great importance attributed to maternal presence, that this topic has not been more frequently included in the type of analyses we have cited here.

SUMMARY

We have presented the results of a factor analysis of child training practices. The data were gathered from mothers in six societies, North India, Mexico, Okinawa, the Philippines, Kenya, and a New England community. A small carefully selected sample of 21–24 mothers was studied in each of these societies. In order to minimize intrasociety variation, the families studied in each country came from the same neighborhood and interacted with each other. The mothers of these samples should not be considered representative samples of their societies, nor are the societies necessarily representative of types of world societies. The factors which were extracted are, however, preponderantly based on differences between mothers (or their practices)

rather than differences in culture-context means.

The correlation matrix of variables was factored by the centroid method and the factors rotated orthogonally. Ten factors were extracted, seven of which were named and interpreted. They are: Responsibility demands on the child, Use of positive affect by mother, Peer aggression and obedience control, Proportion of caretaking of *baby* done by mother, Proportion of caretaking of the *child* done by mother, Control of aggression and disobedience toward parents, and Emotional instability of the mother. Three other factors were discussed tentatively.

Our findings are compared with those of other investigators.

REFERENCES

AHMAVAARA, Y. On the unified factor theory of mind. *Ann. Acad. Sci. Fer., Ser. B*, 1957, **106**, 176.

BALDWIN, A. L., KALHORN, JOAN, & BREESE, F. H. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, **63**(4, Whole No. 299).

LAMBERT, W. W. Interpersonal behavior. In P. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960. Pp. 854–917.

MILTON, G. A. A factor analytic study of child-rearing behaviors. *Child Develpm.*, 1958, **29**, 381–392.

SCHAEFER, E. S., & BELL, R. Q. Structure of attitudes toward child rearing and the family. *J. abnorm. soc. Psychol.*, 1957, **54**, 391–395.

SCHAEFER, E. S., & BELL, R. Q. Development of a parental research instrument. *Child Develpm.*, 1958, **29**, 339–361.

SEWELL, W. H., MUSSEN, P. H., & HARRIS, C. W. Relationships among child training practices. *Amer. sociol. Rev.*, 1955, **20**, 137–148.

TAKALA, M., & TAKALA, ANNIKA. *Child-rearing practices and attitudes as measured by different techniques*. Jyvaskyla, Finland: Centre for Educational Research, 1960.

WHITING, J. W. M., CHILD, I., LAMBERT, W. W., FISHER, ANN, FISHER, J., MINTURN, LEIGH, MARETSKI, HATSUMI, MARETSKI, T., NYDEGGER, CORINNE, NYDEGGER, W., ROMNEY, K., & ROMNEY, ROMAINE. *Field guide for a study of socialization in five cultures*. Cambridge, Mass.: Laboratory of Human Development, 1955.

WHITING, J. W. M., & WHITING, BEATRICE B. Contributions of anthropology to the methods of studying child rearing. In P. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960, Pp. 918–944.

ZUCKERMAN, M., RIBBACK, BEATRICE B., MONASHKIN, I., & NORTON, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, **22**, 165–171.

(Received June 6, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 640-648

SOURCES OF FRUSTRATION AND TARGETS OF AGGRESSION: A CROSS-CULTURAL STUDY[1]

LEIGH MINTURN TRIANDIS
University of Illinois

AND

WILLIAM W. LAMBERT
Cornell University

It has long been considered possible that persons may have more or less characteristic modes of response to frustration—they may either direct the majority of their aggressive responses towards others or towards themselves. In a parallel manner it has also long been considered possible to characterize a person by his habits of attributing the causes of his frustrations to his own actions or to the actions of others. A purpose of the present research is to explore the possibility that small, relatively homogeneous societies may be fruitfully characterized in these ways. Another aim is to investigate the relationships among various types of culturally defined frustrating agents and various types of culturally defined targets of aggressive responses. The results are based on ratings of existing ethnographic reports from a world-wide "sample" of nonliterate societies.

Categories of Analysis

In attempting to classify societies in terms of their methods of aggressive expression we have borrowed from Rosenzweig's (1944) classification of intrapunitive, extrapunitive, and impunitive responses, and have elaborated on this classification, particularly in distinguishing between culturally defined *sources* of frustration and culturally defined *targets* of aggression. Rosenzweig defines three types of aggressive response:

> a. *Extrapunitive responses* are those in which the individual aggressively attributes frustration to external persons or things. The associated emotions are anger and resentment. The cognate psychological mechanism is projection. It should, however, be noted that the extrapunitive response is not always projective. When it proceeds directly from anger it has a very different character from that found when aggression is first inhibited and then only indirectly finds expression in extrapunitive projection. . . .
>
> b. *Intrapunitive responses* are those in which the individual aggressively attributes the frustration to himself. The inturning of aggression is perhaps a consequence of its outward expression. Associated emotions are guilt and remorse. The related psychoanalytic mechanisms are displacement and isolation. . . .
>
> c. Impunitive responses differ from both the extrapunitive and intrapunitive in that aggression does not apparently supply the motivating force; more socially directed or erotic drives are at work. Here the attempt is to avoid blame altogether, whether for others or for oneself, and to gloss over the frustrating situation as though with conciliatory objective. The cognate psychoanalytic mechanism is repression . . . (p. 383).[2]

It should be noticed that Rosenzweig has packed many things into these definitions. For our purposes we have sought in addition: (*a*) to distinguish different classes of apperceived (or culturally defined) frustrators, (*b*) to distinguish between apperceived (or culturally defined) frustrators and apperceived targets of aggression, (*c*) to distinguish among some different classes of apperceived targets of aggression. We have thereby elaborated upon his classification by considering three types of potential external agents who might be culturally defined as frustrators, and three parallel types of agents who might be culturally defined as targets of aggression. The distinctions are made in terms of the "other's" status in relation to "ego," ego being defined as the average or modal adult of a society. The model therefore involves, besides the self, three types of individuals: superiors, peers, and inferiors, who *might* be either culturally defined agents of frustration or culturally defined targets of aggression in any combination. Egos who characteristically define the self as a frustrating agent may be classified as having an intrapunitive outlook. Egos who believe that superiors, peers, or subordinates are frustrating agents have an extrapunitive outlook. Egos who apparently gloss over frustrations, attributing them to none of the classes above, are defined as having an outlook of the impunitive variety.

[1] Research for this paper was facilitated by funds from the Social Science Research Council and the Ford Foundation. Its completion was facilitated by funds from the National Institute of Mental Health.

[2] Quoted from Saul Rosenzweig, "An Outline of Frustration Theory" in *Personality and the Behavior Disorders*, edited by J. McV. Hunt. Copyright 1944 The Ronald Press Company.

In parallel fashion, egos who typically direct some form of aggression against themselves may be said to have intrapunitive *responses*. Egos who typically direct aggression at superiors, peers, subordinates or outgroups (warfare) have extrapunitive responses. Egos who are not aggressive to anyone of the four possible targets are defined as representing impunitive responsiveness.

By stating that the subject "aggressively attributes frustration to" another, himself, or no one, Rosenzweig (1944) assumes that the subject's aggressive response is directed at the apperceived source of frustration. While this is a reasonable hypothesis in many situations, the existence of the phenomena of displaced aggression indicates that it is not always the case that the apperceived source is also the target. A primary purpose of this paper is to explore this very issue: whether or not people of a particular society characteristically direct aggression at those individuals whom they believe to be responsible for frustration—and whether people of societies with an impunitive outlook also exhibit infrequent aggressive responses.

In stating the problem thus we recognize that it can be argued that the aggression, although accompanied by a belief that justifies it, has always actually been engendered by agents other than the ones attacked. Indeed Whiting and Child (1953) assume that displaced aggression is usually rationalized by a justifying belief. The culturally defined beliefs and practices that we are discussing probably have at least partial antecedents in childhood experience, the source of displaced aggression of interest to Whiting and Child. However, the presence of a justifying belief for a type of response makes it difficult to classify the response as displaced in the context of the present analysis. For the purposes of this paper we have considered aggression as direct if it is supported by a belief that the object of aggression is an agent of frustration and as displaced if such a belief does not mediate the aggressive response. Whether or not the aggression is in fact also partially or totally displaced from some other type of aggression rooted in childhood experience is a problem for another analysis.

We also investigate here the relationships that exist in our "sample" among the various types of frustrating agents and among the targets of aggression. Choices of frustrating agents or aggression targets may be mutually exclusive, such that "egos" who "assign" frustration to a particular agent tend not to regard other types of agents as frustrating, and groups who are aggressive to a particular type of individual tend not to be aggressive to other types. If this is the case we would expect to find negative correlations among ratings of categories of agents and targets. If, on the other hand, the choices are independent of each other, we would expect zero correlations among these ratings.

We have used ratings in our analysis that were developed at Cornell, with the exception of the warfare scale. Unfortunately, no measures are available of the belief that subordinates are apperceived as frustrating agents. This aspect of the model must therefore remain untested.

METHOD

Sample

The data of this study consist of ratings made from ethnographic descriptions of nonliterate societies. The total number of societies on which the ratings were made is 60. However, the ethnographies often do not include enough information to enable the raters to make judgments on all the scales, so the actual sample size varies from one portion of the analysis to another. The sample is the same as that described by Lambert, Triandis, and Wolf (1959).

Ratings of Frustrating Agents

All societies face a number of different kinds of frustration. We chose to make ratings about agents believed to be responsible for crop failure or for poor hunting. Most of these ratings were made by graduate students in a research seminar. Ratings on additional tribes were added by scholars employed for that purpose.[3] All ratings were made on four-point scales, "four" representing the belief that a particular agent was always responsible for famine, and "one" representing absence of the belief. Each society was rated by two students, working independently, who subsequently discussed their ratings. If they failed to reach agreement on a particular scale, no rating was entered for that society. The specific ratings follow.

Extrapunitive Outlook—Superiors. It is assumed that "spirits" represent agents who are always viewed as superior to ego. Ratings were made on the importance of the belief that spirits cause damage to the food supply. Separate ratings were made for human male

[3] The authors are grateful to Margery Wolf and Florence Rosenberg for their assistance in rating both frustrating agents and targets of aggression.

spirits, human female spirits, animal spirits, and spirits of indeterminate nature. The highest of these four ratings was used as the measure of extrapunitive outlook towards superiors. The average (independent) interrater reliability for these scales is .87.[4]

Extrapunitive Outlook—Peers. Extrapunitive outlook towards peers was measured by the belief that an adult member of the tribe causes damage to the food supply. Separate ratings were made for women and men and the highest of these ratings was used in the analysis. The average (independent) interrater reliability of these scales is .86.

A second measure of this type of belief was obtained from the ratings on the belief that some outgroup member, usually from some neighboring tribe, is responsible for crop or hunting failure. The independent reliability of this scale is .90.

Extrapunitive Outlook—Overall Measure. An overall measure of extrapunitive outlook was computed by adding, for each society, the scores used in the three preceding scales: the high score for the belief that spirits and tribal members cause famine and the belief that outgroup members cause famine.

Intrapunitive Outlook—"Self-Blame." Self-blame was indexed by the belief that the entire tribe is collectively responsible for damage to the food supply. Since this belief is relatively rare, the scale as used in the analysis represents a measure of presence or absence of the belief. The independent reliability of this scale is .83.

Impunitive Outlook. A measure of impunitive outlook was obtained by adding, for each society, their rating on overall extrapunitiveness and their rating on intrapunitiveness. A low score on this combined measure indicates that members of the society do not blame any of the types of sources considered for failure of the food supply and are therefore defined as impunitive in outlook.

Measurement of Targets of Aggression

These ratings[5] were also made by the students of a research seminar and by paid scholars. The ratings are all on seven-point scales. Two students read the material for each of the tribes. After making independent ratings the students met and discussed those ratings on which there was more than a two point discrepancy, in an effort to reach agreement. If the disagreement between raters was two points or one point, the judgments were averaged and this average score was used for analysis. If the judges could not agree on their ratings to within two points the rating was omitted. A similar procedure was used by the personnel who later made ratings on additional tribes.

Extrapunitive Response—Superiors. This was measured through ratings on the frequency of coercion of gods or spirits. "Coercion" is defined for present purposes as insulting a god or spirit or beating an image of a god or spirit. A rating of seven indicates that such practices occur every day or oftener, and a rating of one indicates that coercion never occurs. The median for this scale falls between—no coercion (1), and the lowest intensity of coercion (2). Therefore the scale as used in the analysis amounts to a measurement of presence or absence of coercion of supernaturals. The independent reliability of this scale is .82.

Extrapunitive Response—Peers. This was measured by ratings on the frequency of verbal aggression among tribal adults. Verbal aggression is defined as nonceremonial quarreling, insulting, and gossiping. Insulting of a nontribal member is not included in this rating. The reliability of this scale is .78.

Extrapunitive Response—Outgroup. This was measured by the frequency with which the tribe engages in warfare. In using these ratings, it must be borne in mind that colonial administrations have put an end to war in most primitive societies. Therefore the warfare ratings often represent the tribal life at a precolonial period of time, whereas the other aggression ratings usually represent the behavior as seen by the ethnographer at the time of the field study. While one may assume that the current aggression of the "ethnographic present" is in some degree representative of the aggressive beliefs and behavior of the society during the period when they were engaged in warfare, one must not ignore the possibility of changes over time in such beliefs and behaviors. Indeed, the suppression of warfare may, in itself, alter other forms of aggressive behavior. The frequency of warfare ratings were collected by Jerome Bruner.

Since Bruner's sample of tribes was smaller than ours and somewhat different, the N for the warfare data is smaller than that involving other scales. The total number of tribes common to both samples is 47. The frequency of warfare was rated on an 11-point scale. In order to make this scale comparable to the seven-point scales of the other ratings on targets of aggression, this scale was arbitrarily condensed to 7 points, by making 0, 1 = 1; 2, 3 = 2; 4 = 3; 5 = 4; 6 = 5; 7, 8 = 6; 9, 10 = 7.

Extrapunitive Response—Inferior. This was measured by the amount of aggression displayed to initiates in the initiation ceremony. This frequency measure, unlike the two preceding scales, is not a measure of aggression over a time period, since initiation ceremonies are seldom held more often than once a year. Instead, the aggression to initiates is measured in terms of the proportion of the total ceremony that is devoted to initiate-directed aggression. The highest rating on this scale means that the entire ceremony is judged as being aggressive. The lowest rating is "no aggression."

In using this variable the question arises as to whether to include tribes that do not have initiation ceremonies and give them a rating of 1 (e.g., "no aggression"), or whether one should limit the analysis to only those tribes where such a ceremony does exist. In the data of this analysis the tribes without ceremonies are included with ratings of 1. Several of the correlations between this and other variables have been computed with the no ceremony tribes omitted; the results remain unchanged with respect to direction or significance. The independent reliability of this scale is .91.

Extrapunitive Response—Overall Measure. A general measure of extrapunitive responsiveness was obtained

[4] All reliability coefficients are Pearson product-moment coefficients.

[5] Although Margery Wolf and Florence Rosenberg rated societies on both agents of frustration and targets of aggression, the students who did the majority of the ratings on frustrating agents were not the same students as those who rated targets of aggression.

by adding, for each society, the scores on frequency of coercion of the gods, frequency of verbal aggression, frequency of warfare, and proportion of time devoted to aggression in initiation ceremonies.

Intrapunitive Response. This was measured by ratings on the frequency of sacrifice. Sacrifice is defined as giving up anything of real to symbolic value to the deity or deities. The independent reliability of this scale is .86.

Impunitive Response. An index of impunitive responsiveness was obtained by adding, for each society, the score on overall extrapunitive responsiveness and the score on intrapunitive responsiveness. A low score in this index indicates that members of the society are generally low on expression of aggression, at least according to the categories entertained in the present conceptual scheme.

These scales, of course, do not represent all possible types of aggression. In the area of agents of frustration we have measured only one type of frustration, threat to the food supply. Our ratings of targets of aggression do not include, for instance, measures of brawls, murder, or suicide, nor do they include ceremonial aggression against any of the societies' members other than initiates. Ratings on frequency of brawls have been made but the variability of the ratings is too low for them to be useful in this analysis. Ratings on frequency of murder and suicide have been made but not for enough of the societies in the sample. In any case, since these are exceptional events involving individuals rather than the group as a whole (ego as defined above), there is some argument in favor of not including them in the present analysis.

We recognize that incomplete coverage of types of aggressive response undoubtedly leads to some error in the classification of societies with respect to overall extrapunitiveness and impunitiveness. However, the assessment of the *total* amount of aggression in a society is a virtually impossible task. Therefore we must simply hope that we have a reasonably representative sample of the aggressive beliefs and responses of the societies studied.

RESULTS

Interrelationships among the Apperceived Frustrating Agents

Table 1 presents the relationships among the various ratings of types of agents of crop and hunting failure. Each case in the contingency tables represents a society. For each analysis the distribution is dichotomized as close to the median as possible. The main finding in Table 1 is that none of the chi square values in this table approaches significance, and most are virtually zero. Evidently beliefs concerning types of famine-producing agents are independent of one another. Unfortunately, our measure of intrapunitive outlook (i.e., the belief that the entire tribe is responsible for failure of the food supply) is rather rare. Although this finding is itself interesting, the infrequency of this belief limits the usefulness of the scale for the purposes of the present analysis.

TABLE 1
RELATIONSHIPS AMONG APPERCEIVED AGENTS OF FAMINE

	Tribal Member			Outgroup			Spirits		
	High	Low	p^a	High	Low	p	High	Low	p
Whole Tribe									
High	0	6	*ns*	4	1	*ns*	5	2	*ns*
Low	8	16		6	15		13	11	
Tribal Member									
High				3	5	*ns*	4	6	*ns*
Low				10	12		14	17	
Outgroup									
High							8	6	*ns*
Low							9	9	

[a] p is based on χ^2 values. All probabilities are based on two-tailed tests.

Interrelationships among Targets of Aggression

Table 2 presents the relationships among the different targets of aggression. The most striking pattern that emerges from this table is the consistent presence of a fairly high negative relationship between frequency of verbal aggression to peers and the other measures of aggressive response. The probability of the negative association between verbal aggression and frequency of sacrifice or of coercion of the gods is less than 5% and the probability of the relationship with aggression in initiation ceremonies is less than 10%.

While the relation between verbal aggression and frequency of warfare is not significant when societies without any warfare are included in the analysis, the relationship becomes negative and significant at the 5% level, if only societies with some token of warfare are considered. Since total absence of warfare may represent either lack of aggressive intent or lack of opportunity due to isolation, lack of resources to engage in warfare, or colonial control, it seems reasonable to consider only those tribes with at least some frequency of warfare greater than zero in making this comparison.

The other significant relationship among targets of aggression is the positive association between frequency of sacrifice and amount of aggression to initiates.

TABLE 2
RELATIONSHIPS AMONG RECIPIENTS OF AGGRESSION

	Frequency of Verbal Aggression			Amount of Aggression to Initiates			Frequency of Coercion of the Gods			Frequency of Warfare			Frequency of Warfare[a]		
	High	Low	p[b]	High	Low	p	High	Low	p	High	Low	p	High	Low	p
Frequency of Sacrifice															
High	7	12	.02	16	7	.02	11	9	*ns*	11	7	*ns*			
Low	17	6		6	13		10	14		7	13				
Frequency of Verbal Aggression															
High				7	13	.10	3	12	.01	4	10	*ns*	4	10	.05
Low				14	9		13	7		9	9		9	4	
Amount of Aggression to Initiates															
High							10	12	*ns*	13	6	.10			
Low							15	9		7	12				
Frequency of Coercion of Gods												*ns*			
High										6	12				
Low										9	6				

[a] Analysis includes only those societies having token warfare or more.
[b] p is based on exact χ^2 values.

Relationships between Apperceived Frustrating Agents and the Target of Aggression

Table 3 shows the relationships among assumed agents of famine and targets of aggression. The intrapunitive belief that the entire tribe is responsible for famine is not, as might be expected, related to the intrapunitive response of sacrifice, but is most positively associated with frequency of coercion of the gods ($p < .10$).

In the realm of extrapunitiveness, we find that the belief that some tribal member is responsible for failure of the food supply is positively related to the extrapunitive response of aggression to initiates ($p < .05$), but not to verbal aggression to peers. It is negatively associated with frequency of warfare ($p = .02$). The belief that some outgroup member is the culprit is not, as might be expected, related to frequency of warfare, but is positively associated with coercion of the gods ($p < .10$). The overall index of extrapunitive outlook is also positively associated with frequency of coercion of the gods, while the index of impunitive outlook is negatively associated with this variable.

Aside from the associations just mentioned, the indices of overall extrapunitive response and of impunitive response do not relate to any other measures of frustrating agents. It is particularly disappointing that the index of extrapunitive response is not highly associated with the index of extrapunitive outlook and that the measures of impunitive outlook and impunitive response are not related to each other. We also find that the measure of frequency of warfare does not seem to be a particularly sensitive measure of extrapunitive responsiveness. In view of the archaic nature of the frequency of warfare ratings, we re-computed the indices of extrapunitive and impunitive responsiveness, omitting this measure, and including in the analysis only those societies where some outside authority, such as a colonial administration, had passed laws against warfare.[6] For such societies one may assume that the extrapunitive outlet of warfare is impossible, or extremely limited. The results of analyses with these indices are presented in Table 4. Also presented are the relationships between frequency of warfare and agents of famine for those societies where at least token warfare is present.

The elimination of tribes without warfare does not substantially alter the relationships between frequency of warfare and agents of famine. Frequency of warfare for this re-

[6] Forty-three of the societies in our sample have externally imposed laws against war, 10 societies have no such laws, and 4 were rated as "don't knows." No ratings on this matter have been made for the remaining societies. These ratings were made in a manner similar to the other ratings.

TABLE 3

RELATIONSHIPS AMONG ASSUMED AGENTS OF FRUSTRATION AND TARGETS OF AGGRESSION

Apperceived Agents of Famine	Frequency of Sacrifice			Frequency of Verbal Aggression			Amount of Aggression to Initiates			Targets of Aggression: Frequency of Warfare			Targets of Aggression: Frequency of Coercion of the Gods			Targets of Aggression: Overall Extrapunitive Responses[a]			Targets of Aggression: Impunitive Responses[b]		
	High	Low	p[c]	High	Low	p	High	Low	p	High	Low	p	High	Low	p	High	Low	p	High	Low	p
Whole Tribe																					
High	4	2		3	3	ns	1	5	ns	2	1	ns	5	1	.10	3	0	ns	0	3	ns
Low	11	13	ns	8	11		13	10		7	10		7	12		5	7		8	5	
Tribal Member																					
High	8	5	ns	7	8	ns	6	4	.05	3	10	.02	6	5	ns	4	4	ns	4	4	ns
Low	17	16		10	12		12	20		13	7		13	16		7	8		7	7	
Outgroup																					
High	8	6	ns	7	7	ns	9	6	ns	6	4	ns	8	5	.10	7	4	ns	3	6	ns
Low	7	10		8	7		7	8		5	9		4	11		3	6		7	4	
Spirits																					
High	11	12	ns	9	7	ns	9	10	ns	7	7	ns	8	9	ns	5	5	ns	4	6	ns
Low	10	9		9	9		13	10		6	7		10	10		5	6		6	4	
Overall Extrapunitive Outlook[d]																					
High	10	10	ns	8	9	ns	9	10	ns	5	9	ns	9	8	.10	6	6	ns	6	7	ns
Low	3	8		5	5		4	7		7	4		3	10		3	4		4	2	
Impunitive Outlook[e]																					
High	3	5	ns	4	4	ns	2	5	ns	4	7	ns	4	11	.05	7	4	ns	7	4	ns
Low	9	6		6	6		9	7		4	4		5	2		2	3		1	4	

[a] Verbal Aggression + Aggression to Initiates + Coercion of Gods + Warfare.
[b] Verbal Aggression + Aggression to Initiates + Coercion of Gods + Warfare + Sacrifice.
[c] p is based on exact χ^2 values.
[d] Tribal Member + Outgroup + Spirits.
[e] Tribal Member + Outgroup + Spirits + Whole Tribe.

TABLE 4

RELATIONSHIPS AMONG AGENTS OF FAMINE AND TARGETS OF AGGRESSION FOR SOME SUBSAMPLES OF THE TOTAL *N*

Assumed Agents of Famine	Targets of Aggression: Frequency of Warfare[a]			Targets of Aggression: Overall Extrapunitive Response[b]			Targets of Aggression: Impunitive Response[b]		
	High	Low	p	High	Low	p	High	Low	p
Whole Tribe									
High	1	6	ns	3	2	ns	3	1	ns
Low	1	6		6	7		5	9	
Tribal Member									
High	3	13	.02	6	3	ns	5	4	ns
Low	6	3		6	10		11	5	
Outgroup									
High	5	5	ns	7	4	ns	5	3	ns
Low	1	7		3	6		4	5	
Spirits									
High	5	6	ns	6	7	ns	6	7	ns
Low	5	4		6	7		4	5	
Overall Extrapunitive Outlook									
High	4	7	ns	9	5	.20	6	8	ns
Low	5	3		1	3		3	2	
Impunitive Outlook									
High	3	3	ns	4	6	ns	6	3	.10
Low	4	3		4	2		1	4	

[a] Analysis includes only those societies having token warfare or more.
[b] Index computed only for those societies with externally imposed laws against warfare.

stricted sample is still negatively related to the belief that some tribal member is responsible, but is still not related to the index of the belief that the outgroup is a frustrating agent.

The elimination of the warfare ratings from the indices of extrapunitive and impunitive response does alter the picture somewhat for these indices. For this sample, impunitive responsiveness is positively related to the index of impunitive outlook ($p < .10$). There is also a tendency for overall extrapunitive outlook to be positively associated with overall extrapunitive responsiveness, although this trend does not reach any conventional level of significance.

DISCUSSION

Interrelationships among Agents of Frustration

Beliefs concerning the type of agents responsible for famine are evidently relatively independent of one another, since no significant relationships appear in Table 1. Societies may be measured with respect to the prevalence of beliefs in various frustrating agents in this area, but they usually cannot be typed as favoring exclusively one or another form of belief.

Interrelationships among Targets of Aggression

As we have said, the most consistent pattern in Table 2 is the consistency of negative relationships between the extrapunitive response of verbal aggression and all other measures of aggressive response. This can perhaps be best explained as an inverse relationship between *spontaneous* and *structured* aggression. Verbal aggression is the only measure of aggression that takes place outside of a structured context: it may occur anywhere, at any time. In contrast, our measures of sacrifice, coercion of the gods, and aggression to initiates all occur in the course of some ceremony, the timing of which may not be decided by the aggressor. The timing of attacks in war is similarly beyond individual control, depending upon strategic considerations or the attack of the enemy. This difference, note, is not a distinction between direct and displaced aggression, since any of the four directions of aggressive response might be instigated either by frustration from the agent against whom the aggression is directed or by frustration from some other agent. Nor can it be assumed that verbal aggression necessarily follows immediately upon the frustrating event, since both gossip and direct insults may well be delayed actions. But delayed or not, the timing and form of the verbal aggression may be chosen by the aggressor and presumably occurs in the presence of hostile emotions, whereas the ceremonial aggression and warfare usually occur at a predetermined time when, at least in the case of ceremonial aggression, the aggressor may or may not experience hostile emotions. It would seem that one can distinguish two kinds of societies, one in which people express their aggression spontaneously in unstructured situations and another in which they devise more structured means for aggressive outlet. These data provide some justification for Rosenzweig's (1944) contention, quoted above, that: "when it (aggression) proceeds directly from anger it has a very different character from that found when aggression is first inhibited and then only indirectly finds expression in extrapunitive projection" (p. 383).

The other significant relationship among recipients of aggression is the positive association between frequency of sacrifice and amount of aggression toward initiates. The reason for this relationship is not clear. It is possible, of course, that aggression directed toward initiates is aggression directed toward the children of ego, and, therefore, toward "his" extended ego. Further research on the antecedents or concomitants of both practices is necessary to determine the connection between them.

Interrelationships between Agents of Frustration and Targets of Aggression

The relationships presented in Table 3 are designed to answer the primary question of this research: whether or not societies typically direct their aggression to those agents whom they believe to be frustrating or whether they displace aggression to (possibly) innocent individuals.

Looking at the measure of intrapunitive outlook one sees that the intrapunitive belief that the entire tribe is responsible for famine is not related to the intrapunitive response of sacrifice. One cannot regard this test as conclusive because of the low frequency with

which our index of intrapunitive outlook is counted as present. However, these data fail to demonstrate a relationship between intrapunitive outlook and intrapunitive response.

There is, however, some evidence for a positive association between extrapunitive outlook and extrapunitive response. None of the measures of extrapunitive outlook (i.e., beliefs that tribal members, outgroup members, or spirits are responsible for famine) are related to either the intrapunitive response of sacrifice or to the index of impunitive responsiveness. If such aggression takes the form of fairly direct retaliation, we would expect that the belief that a tribal member is responsible for famine would be positively associated with aggression toward members of the tribe. We have two ratings of such aggression, verbal aggression and aggression to initiates. Belief in the guilt of tribal members is positively related to one of these responses, aggression against initiates ($p < .05$). This is, however, the only relationship in which aggression is directed against the kind of agent believed to be responsible for the frustration. The belief that the outgroup causes famine is not significantly related to frequency of warfare, nor is belief that the spirits are frustrating agents related to coercion of the gods and spirits.

We do find, however, that the extrapunitive belief that the outgroup is responsible for famine, and the index of overall extrapunitive outlook, are positively associated with the coercion of the gods and spirits. This might indicate that an extrapunitive belief may result in an extrapunitive response, but one which is directed against some target other than the one believed to be frustrating. The deities seem to be a popular target of aggression that is engendered from other sources. This being the case, it is somewhat surprising that coercion of the deities is not related to the belief that the spirits themselves are responsible. The measure of spirits as agents of famine seems to be a somewhat insensitive index, since it is not significantly associated with any other measure. In order to check on the possibility that it is this particular measure of aggressiveness of the deities which is responsible for the lack of relationship, we plotted the relationship between coercion of the gods and a general rating of the frequency with which the gods are aggressive in all areas of their behavior.[7] The probability of this positive relationship is less than 2%. Frustration by the deities is, then, associated with aggression against them, when this broader measure is used.

The relationship between aggression by the gods and aggression to them may be the explanation for the positive correlation between the belief that the whole tribe is responsible for famine and frequency of coercion of the gods ($p < .10$). This is the only example of a response that is not accompanied by a belief falling at least within that category of response, according to Rosenzweig's (1944) classification. This association can perhaps be explained by the nature of the deviation that causes the failure of the food supply. Frequently the crime committed by the tribe is an offense against the gods. The group fails, in some way, to pay homage to the gods, either by some sacrilegious act or through failure to perform some sacrifice or ceremoney. In retaliation the gods send famine. Therefore, although the group is responsible, the gods are also responsible for the misfortune. Viewed in this fashion, the relationship is an extrapunitive-extrapunitive one. The society retaliates against their hostile gods by being in turn hostile to them.

There is one negative relationship between an extrapunitive belief and an extrapunitive response, that between the belief that a member of the society causes famine and frequency of warfare. It seems reasonable that if warfare is a direct response to frustration, it would be unlikely to occur when some member of the ingroup is perceived as the frustrating agent. This argument would carry more weight, however, if warfare were positively related to the belief that the outgroup causes famine.

The general positive association between the indices of impunitive outlook and impunitive response presented in Table 4 is perhaps the most interesting finding of this research. Certain societies are apparently nonaggressive both in outlook and response. In future work we hope to examine other characteristics of these impunitive societies and their aggressive

[7] This rating was developed in the same manner as those ratings presented in this text. The frequency of aggression of the gods was measured against the baseline of their total behavior (see Lambert, Triandis, & Wolf, 1959).

counterparts that may account for this variation in aggressive tendencies.

It would seem, in summary, that the types of aggressiveness of Rosenzweig's (1944) classification at the societal level, at least, are better viewed as relatively independent variables rather than mutually exclusive categories. Beliefs about apperceived *agents* of frustration are independent of each other, and the negative associations among types of aggressive *responses* are generally between spontaneous vs. ceremonialized aggression rather than one type of aggressive target and another. Furthermore, for the most part, intrapunitive, extrapunitive, and impunitive beliefs are independent of responses in other categories (e.g., extrapunitive *beliefs* are unrelated to either impunitive or intrapunitive *responses*). Rosenzweig's implied contention that the aggressive response is typically directed at the apperceived agent of frustration is upheld for extrapunitiveness and impunitiveness. Intrapunitive outlook was found to be unrelated to intrapunitive responsiveness. This lack of relationship may, however, be due to the relative rarity in the cultures studied of the bases for our measure of intrapunitive belief.

SUMMARY

This research represents a cross-cultural investigation of relationships between beliefs about people thought to be responsible for damage to the food supply and types of people against whom the members of these societies direct their aggressive responses. Rosenzweig's (1944) classification of intrapunitive, extrapunitive, and impunitive types of aggressiveness was employed and elaborated upon.

Ratings were made on the basis of ethnographic material on a world-wide sample of 60 primitive societies. The ratings included judgments of intrapunitive, extrapunitive, and impunitive beliefs about famine-producing agents and intrapunitive, extrapunitive, and impunitive types of response. Extrapunitive beliefs and responses were further distinguished in terms of whether the frustrating agents or targets of aggression were superior, equal to, or inferior in status to the typical adult male of the society.

No evidence was found for intrapunitive responsiveness as a reaction to intrapunitive outlook. Instead, intrapunitive outlook was found to be positively associated with the extrapunitive response of coercion of the gods.

Extrapunitive outlook was generally found to be associated with extrapunitive responsiveness. Impunitive outlook was found to be associated with impunitive responsiveness.

The data are discussed in terms of displacement, projection, and Rosenzweig's classification.

REFERENCES

LAMBERT, W. W., TRIANDIS, LEIGH M., & WOLF, MARGERY. Some correlates of beliefs in the malevolence and benevolence of supernatural beings: A cross-societal study. *J. abnorm. soc. Psychol.*, 1959, **58**, 162–169.

ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 379–388.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

(Received June 6, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 649-658

THE PERPETUATION OF AN ARBITRARY TRADITION THROUGH SEVERAL GENERATIONS OF A LABORATORY MICROCULTURE

ROBERT C. JACOBS[1] AND DONALD T. CAMPBELL
Northwestern University

ACCORDING to Sumner (1906) "the mores can make anything right," and he went on to illustrate with a chapter on sacral harlotry and child sacrifice. According to Tarde (1903) "imitation is the key to the social mystery," "society is imitation," "social man is a veritable somnambulist." Such quotations indicate the late 19th century awareness of the power of culture to perpetuate arbitrary beliefs. Less drastically expressed, some such perspective permeates present day social science, although tempered somewhat by functionalism in anthropology and sociology. In social psychology, for example, cultural tradition tends to be invoked primarily as an "explanation" for social evils, prejudice, resistance to ameliorative social change, and the like. The current emphasis upon conformity reinforces this view, in spite of Asch's (1952) demurrer. Yet there are, no doubt, restraints upon a complete arbitrariness to culture, particularly with regard to beliefs that lie within people's direct range of observation. Thus arbitrary and erroneous superstitions about the shape of fishhooks probably would not long survive the systematic drift pressures resulting from continual minor variations some of which are in the direction of noticeably more effective form. Here lies the problem in the theory of culture for which a laboratory analogue was sought.

Since Sherif's (1936) classic studies on the formation of "social norms" in laboratory groups, there have been sporadic efforts to bring the process of cultural transmission into the laboratory. While much small group research might be so interpreted, the present study was in particular inspired by Sherif's studies, by Rose and Felton's (1955) "experimental histories of culture," and by the preliminary report of experiments on the evolution of "microcultures" by Gerard, Kluckhohn, and Rapoport (1956). In the latter two studies and in the present one, there is an effort to demonstrate a perpetuation of "cultural" characteristics that transcends the replacement of individual persons. In the present study confederates have been employed to establish an extreme cultural norm, the inculcation and survival of which is then studied as the confederates are one by one taken out of the group, naive new members gradually introduced, who then unwittingly become the further transmitters of the belief to still newer entrants.

The specific culture trait and the ecology of the group relative to such traits should obviously affect the success in the indoctrination of each new generation, and the rate of erosion of the tradition by innovation. In the Rose and Felton (1955) study, group interpretations of Rorschach cards were employed, a material offering considerable latitude for cultural arbitrariness and idiosyncrasy. In the Gerard, Kluckhohn, and Rapoport (1956) study, the groups were continually faced with new instances of a very difficult puzzle series, and were continually confronted with evidences of the lack of perfection of their traditional solutions, a situation encouraging the evolution of more adaptive norms, and unfavorable to the perpetuation of pure superstition. For the present study, a task was sought which would allow as much latitude for cultural arbitrariness as the Rose and Felton situation, and which would, in addition, make possible the ready quantification of effects. The original Sherif (1936) autokinetic movement situation was judged to provide this: if a person views a pinpoint of light in an otherwise totally blacked-out room, the light soon appears to move. The illusion is a strong one for most persons, and is apparent even to those who know the light to be stationary. As Sherif has shown, naive respondents are very suggestible as to the degree of movement seen, and laboratory group norms on the amount of movement are rapidly established without the respondents

[1] Now at the University of Colorado Medical School.

being aware of the fictitiousness of their judgments.

So labile is the autokinetic experience or at least the translation of it into judgments of linear extent, that one reading the reports of studies employing it might expect that an arbitrary group norm once established would be passed on indefinitely without diminution; that once well indoctrinated, the naive group members would become as rigid and reliable spokesmen for the norm as were the confederates who preceded them; that each new generation would unwittingly become a part of a self-perpetuating cultural conspiracy propagating superstition and falsehood.

But the autokinetic experience is not completely labile. The very fact that sophisticated observers under strong "suggestion" to see the light remain fixed still perceive movement, shows this. Haggard and Rose (1944), in attempting to condition perceived movement on a right-left dimension, found for most respondents a strong rightward bias. In preliminaries to the present experiment, we learned that for our setting, respondents reporting alone produced mean individual judgments over 30 trials of from .64 to 6.67 inches, and with individual judgments rarely if ever approaching the arbitrary cultural norm of 15 to 16 inches which the confederates in our study provide. It is the presence of this potential natural norm that makes our inculcated cultural norm truly "arbitrary," and provides a counterpressure to the cultural tradition in its transmission across the experimental generations.

METHOD

Respondents

The 175 respondents who took part in this experiment were students, enrolled in introductory social sciences courses. They were unsophisticated concerning the autokinetic phenomenon and were unadvised as to the nature of the experiment previous to the time of serving in it. The respondents were assigned randomly to the six conditions involved. The number of respondents participating per group is shown in Table 1.

Materials, Apparatus, and Common Procedure

The experiment took place in a completely darkened, windowless, fan-ventilated room. The respondents were seated in a row 8 feet from a box designed to emit a small pinpoint of light. A 1 rpm induction motor was attached in series with the lamp, and a single switch controlled both. The motor was used solely as an auditory effect supporting the illusion of movement.

All respondents were blindfolded when brought into the experimental situation and were asked to put their blindfolds on whenever the door was opened to allow old respondents to leave and new ones to enter. This gave the respondents little if any knowledge of the size and arrangement of the experimental room. The individual respondents received the following instructions:

> This is a movement experiment. It is designed to test the students' visual perceptions with respect to space. In the next room is an apparatus that will project a small light. A few seconds after we begin, the light will move. It may move in a wavy motion or pattern or it may follow a recognizable course. However, your specific task will pertain merely to judging the distance that the light moves from the time that it appears until it is turned off. All judgments should be made in inches along a straight line connecting the starting point and the ending point on each trial.

Respondents going into group conditions received these additional instructions:

> You are going to join a student (or number of students) who are already serving in this experiment. . . . You will be seated in the seat on the far left. To simplify recording procedure I will ask you to report your judgments beginning with the person on the far right and ending with you on the far left. When I stop you the person on the far right will leave, each of you inside will move right one seat, and a new person will be brought in.

Each trial was 5 seconds in length and after each block of 30 trials a ventilation fan was turned on by the experimenter. During this time in the group conditions, the "oldest" member of the group was taken out, other members moved to the right one seat, and a new member introduced. In the solitary conditions, respondents were told that ventilation took place as a health requirement. The ventilation period was timed to take approximately the same amount of time that was necessary to give instructions to a new respondent and bring him in in the other conditions.

Experimental Conditions

Table 1 shows the main feature of the six treatment conditions. (These have been given designations sum-

TABLE 1
EXPERIMENTAL PARADIGM

Condition	Size of Group	Number of Confederates	Trials per Generation	Number of Generations	Generations per Respondent	Number of Replications	Number of Naive Respondents
C-1-0	1	0	30	—	4	24	24
C-3-0	3	0	30	10	3	3	30
X-2-1	2	1	30	9	2	3	27
X-3-2	3	2	30	10	3	3	30
X-4-3	4	3	30	11	4	3	33
X-3-1	3	1	30	9	3	3	29

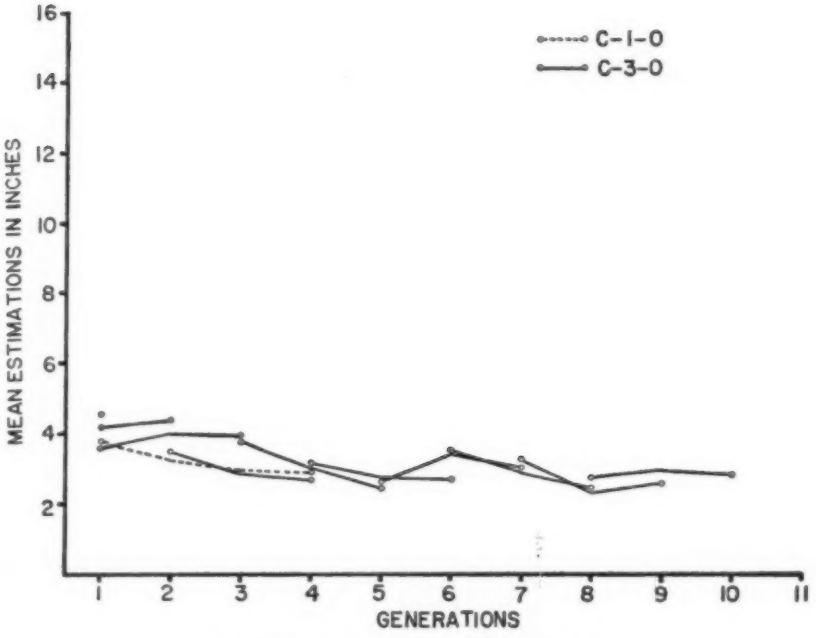


FIG. 1. Mean judgments of the control groups.

marizing the group's construction. C and X stand for control and experimental. The first digit indicates the size of the group, while the second digit indicates the number of confederates present in the first generation.) There are two control conditions. In C-1-0 each respondent judged the movement of the light in solitude for four periods (called generations for the group conditions) of 30 judgments each. In C-3-0 respondents were run in groups of three, replacing one each generation (or 30 trials), for a total of nine generations.

A major dimension of experimental variation was the number of "culture-bearing" confederates with which the first naive respondent found himself placed. In experimental groups X-2-1, X-3-2, and X-4-3, the initial generation consisted of a solitary naive respondent sitting to the left of one, two, or three confederates who gave their judgments before he did, and who had been instructed to give judgments between 15 and 16 inches. The experimental variation in number of confederates was expected to produce cultural traditions of increasing strength. Confederates were recruited from the same subject pool as were the naive respondents. All were pledged to secrecy, with apparent success.

Another mode of manipulating cultural strength was explored by adding X-3-1, a group of two naive respondents led by one confederate. Thus C-3-0, X-3-1, and X-3-2 provide a set of three-person groups of 0, 1, and 2 confederates in the first generation.

The confederates who were present in the starting condition were removed one at a time, after each round of 30 judgments. Thus the groups with more confederates present in the first generation likewise had some stooging present for more generations. The shifting of members one chair to the right each generation, plus the rule of responding in turn from right to left, insured that on each trial a confederate or the eldest member of the group spoke first.

Each of the groups was replicated three times, while 24 respondents were run in the solitary control condition. The conditions were given in a counterbalanced order insuring that the differences in respondent selection from one part of the term to the other were kept independent of treatments.

RESULTS

Control Treatments

Figure 1 shows the results for the two control conditions. It will be noted that for C-1-0 the mean of the 24 respondents starts at 3.80 and decreases steadily generation by generation to a value of 2.90. This decrease is highly significant statistically, being found in the individual records of 21 of the 24 respondents, for which, by a two-tailed sign test, $p < .0002$.

The three-person group control, C-3-0, shows very comparable mean levels, and likewise a general decline. The effect is not so clearly significant and is beclouded by the mutual

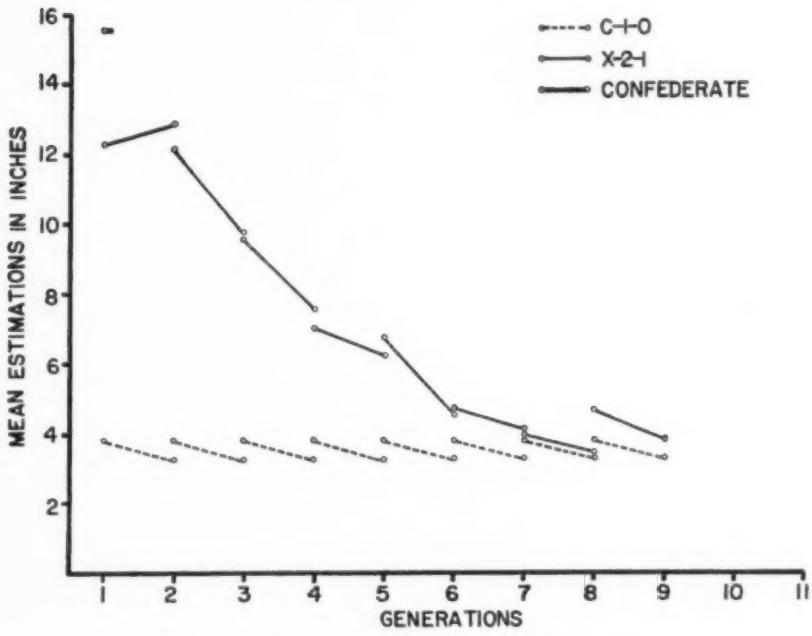


FIG. 2. Transmission of arbitrary norm in two-person groups.

influence that the group situation introduces. In the three replications, two groups declined, while one started abnormally low and increased. Comparing Generation 1 with the last generation with three persons present, Number 8, average judgment changes were 3.93–2.90, 6.72–2.03, and 1.72–2.49. Notice that in spite of the idiosyncratic norms of Generation 1, the three replications became more similar as the process of successive replacement went on. Of the 27 respondents present in more than one generation, 16 showed declines between their first and last generations.

In graphing C-3-0, the data have been averaged in a different fashion from the foregoing. To show the successive replacement and overlapping generational character, the careers of each successive replacement have been plotted separately. The plotted points represent the average of the three replications; that is, each plotted point represents the average judgment of three individual respondents, each taking the corresponding role in the three separate groups replicating this control treatment.

The two types of control established such similar reference standards that they cannot both be used conveniently as visual base lines in the graphing of the results of the experimental treatments. For this reason, in Figures 2, 3, and 4, only C-1-0 has been plotted. This plotting has been done in such a manner as to provide a parallel for each new starting generation in the experimental group. Thus in Figure 2, where each respondent stays only two generations, the first two generations of data from C-1-0 have been repeatedly plotted. In Figure 4, all four generations of C-1-0 are needed as a comparison base.

Transmission of Arbitrary Norms over Total Replacement of Indoctrinator

Figures 2, 3, 4, and 5 present the transmission results from X-2-1, X-3-2, X-4-3, and X-3-1. The first question that we ask of the data is whether or not the naive respondents, once indoctrinated by the confederates, have

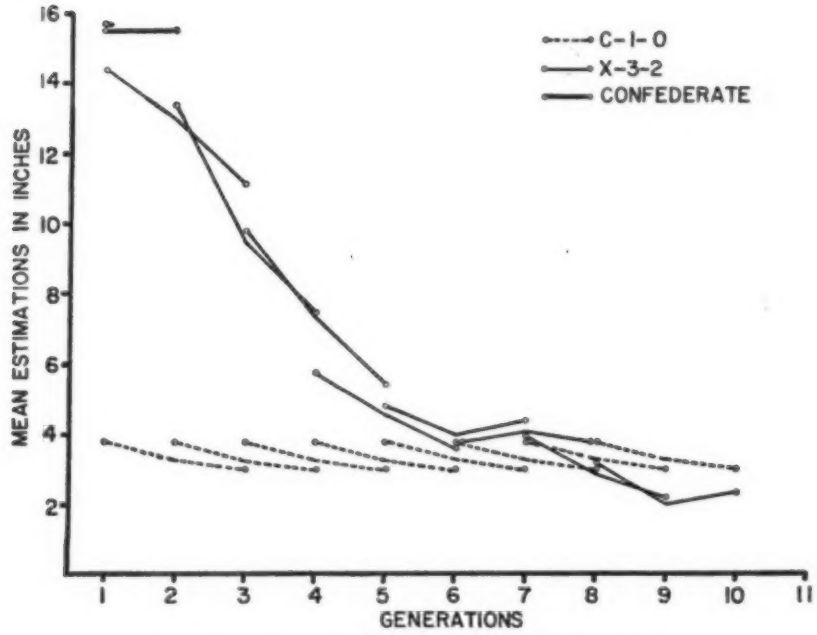


FIG. 3. Transmission of arbitrary norm in three-person groups.

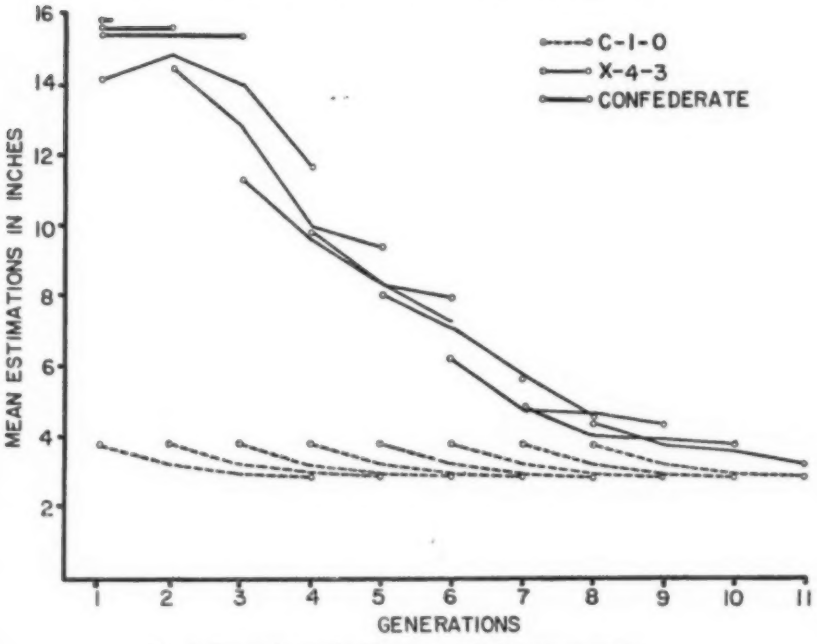


FIG. 4. Transmission of arbirtary norm in four-person groups.

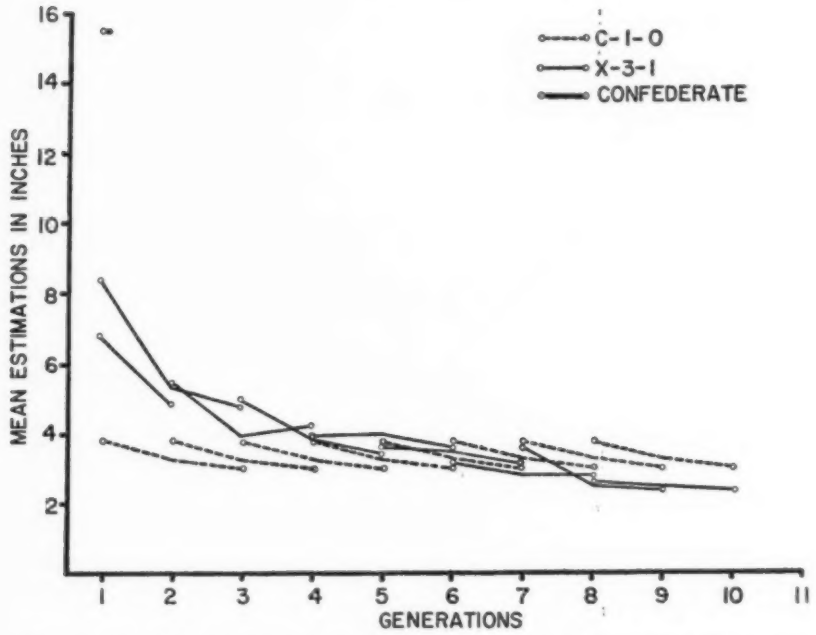


FIG. 5. Transmission of arbitrary norm in a three-person group initiated by one confederate and two naive respondents.

themselves transmitted any vestiges of the arbitrary cultural norm once the original indoctrinators have passed on. To test this, one can examine the judgments of the first generation of respondents to judge without any confederates present. For X-2-1, X-3-2, and X-4-3 these values are 12.07, 9.79, and 9.82. Each of these differs significantly ($p < .01$) from the C-1-0 control group by t test, when the mean of the 3 experimental respondents is compared with the mean of the 24 control respondents. For X-3-1, the mean is 5.34, not significantly different from the C-1-0 value of 3.80.

By pooling groups X-2-1, X-3-2, and X-4-3, we can compare to the 24 C-1-0 respondents 9 experimental respondents newly introduced at each of several generations beyond the final confederate. For the first such generation, $M = 10.57$, $t = 7.25$, $p < .0001$; for the second, $M = 7.80$, $t = 3.74$, $p < .0001$; for the third, $M = 6.06$, $t = 4.02$, $p < .0001$; for the fourth, $M = 5.08$, $t = 2.04$, $p < .03$; for the fifth, $M = 4.37$, $t = 1.31$, $p < .10$. Thus the arbitrary norm is transmitted in some degree up through the fourth and perhaps the fifth generation beyond the last confederate. By the sixth it has entirely disappeared, the mean of 3.60 being slightly below the control value of 3.80.

Differences in Strength of Induced Traditions

The experimental comparisons were introduced in the expectation that induced cultures of differing strengths and persistence would be produced. In fact, there was some expectation that the strongest conditions might produce an arbitrary culture which would persist without apparent diminution. No condition produced a culture of any such strength, and among the three experimental groups starting with one naive respondent, the number of confederates in the groups had little if any clear-cut effect. In the first generation, with all confederates present, the means for the naive respondents were, for X-2-1, X-3-2, and X-4-3, respectively, 12.23, 14.37, and 14.13. X-2-1 was expected to be lower than the others.

The difference between 12.23 and 14.37 is significant ($t = 2.16$, $df = 4$, $p < .05$, one-tailed test). The absence of a difference between X-3-2 and X-4-3 is against our expectation. The literature on this point is confusing. Asch (1951) finds almost no effect from a single confederate, whereas we, in X-2-1, and Goldberg (1954) and Kidd (1958) find strong effects from a single other. Asch finds the effect to increase strikingly when the confederates are increased to 2 and again to 3, whereas 4, 8, and 16 others have little or no greater influence than 3. Goldberg and Kidd find no significant gains between 1 and 3 (that is, between 2- and 4-person groups) in spite of large numbers of cases.

A comparison of the overall strength of the induced cultures is difficult because of the fact that in the larger groups not only are the culture-inducing confederates more numerous, but they are also retained for more generations. Thus it would seem unfair to compare all groups at the fourth generation from the beginning, since X-4-3 would have had a confederate in the immediately preceding generation, while X-2-1 would have been without a confederate for three generations. Perhaps more appropriate is a comparison of strength in the first generation subsequent to the removal of the last confederate. The values are, respectively, 12.09, 9.79, and 9.82. The smallest group, X-2-1, seems strongest at this stage although the differences do not approach significance. But this comparison is not exactly fair, either, because for X-2-1 the confederate, when present, dominated the culture, while for X-4-3 the final confederate had his influence diluted by the vocal presence of two not completely loyal naive respondents. By the fifth postconfederate generation, these three values are 4.69, 3.90, and 4.55. Lacking both striking differences and an optimal comparison, it seems fair to conclude that no differences in effective cultural strength have been achieved by thus manipulating group size.

One manipulation did indeed provide indoctrinations of varying magnitude. The naive respondents of X-3-1 were significantly different from C-1-0 only in the first generation, that is, with the confederate present ($t = 3.88$, $p < .01$). But even at this stage the naive respondents of X-3-1 were significantly less indoctrinated than those of X-3-2 at Generation 2, when it likewise consisted of but one confederate and two naive persons ($t = 4.69$, $p < .001$, 10 df). Thus manipulating the ratio of confederates to naive judges has little effect when achieved through adding more confederates, but has great effect when achieved through manipulating the number of naive respondents. As Asch (1951) has noted, this adding of a "true partner" seems to have an effect disproportionally greater than would be expected from the simple fact of diluting the majority. This dimension has not been systemically explored, however. In the present setting it would be interesting to compare groups of varying size in which the naive-to-confederate ratio was held constant, as through a group consisting of two confederates and two naive respondents, to be compared with X-2-1, etc.

Judgments as a Compromise between Own and Other's Observations

The deterioration of the cultural norm occurs, in part, because each naive respondent makes judgments somewhat lower, somewhat closer to the natural norm, than do the confederates or elder citizens serving as his mentors. Of 36 naive respondents serving for the first time in X-2-1, X-3-2, and X-4-3 during the first four generations (where some significant strength to the cultural norm was still present) 34 made mean judgments lower than the average of their mentors (for which, $p < .0000001$). This clear-cut fact we take as supporting the interpretation that each judgment represents a pooling of the person's own observations (the value for which we infer from the control group) with the reported observations of others. In the resulting compromise the person's own observations are not given much weight, but they are given some.

In the first generation of the experimental groups, the conditions are simple enough so that we can make some estimates of this relative weighting of own observations and other's reports. For the first naive respondent in X-2-1, the other's (confederate's) report averages 15.52. The own observation is inferred from C-1 to be 3.80. If the naive respondent weighted his own observation equally with the report of the confederate, his judgment would

fall halfway between these values, at 9.66. Instead, the judgment is 12.23, much closer to the confederate's. Rather than regarding the confederate as an equally good observer as himself, the respondent has weighed the confederate's opinion 2.6 times as heavily as his own. The formula for the weighting ratio is (12.23–3.80)/(15.52–12.23). Assuming each confederate to be equally weighted, and solving for the weight given a single confederate's opinion, we get values of 4.6 for X-3-2 and 2.5 for X-4-3. Since these ratios are based upon the performances of but three naive respondents each, they are quite unstable, and are provided here primarily for illustrative purposes. Similar computations for the later generations have been omitted because of the difficulty of taking into account the reciprocal effect of new respondents upon the older naive respondents acting as indoctrinators. The first generation situation is computable just because the confederates were unresponsive to the judgments of their fellows.

Whatever the weighting ratio, as long as some weight is given own observations, and as long as own observations deviate in a consistent direction from the cultural norm, the arbitrary culture is doomed to deterioration eventually, at least in the limited sanction system of this laboratory. While the question has not been stated in this form, the relative respect for own observations over the reports of others is much higher in less ambiguous situations, as Asch (1951) has shown. It is easy to understand how the ambiguities of the autokinetic situation lead to lack of confidence in own opinion, but it is perhaps an indication of human trust that this same ambiguity does not lead to a proportional lack of confidence in the reports of others.

Forgetting and Reciprocal Influence

If the previous factor of compromise weighting were all that was involved, the deterioration in culture as graphed in Figures 2, 3, and 4 might have been of a stair-step fashion, with each new inductee starting lower than his indoctrinator, but once started, holding to his initial level through the several generations of his life in the group. As can be seen from the graphs, this is not the case. Instead, the respondent is less loyal to the old culture the longer he stays in the group. Of the 36 naive respondents introduced in the first four generations of X-2-1, X-3-2, and X-4-3, 31 made lower mean judgments on their last session than on their first (for which, $p < .00001$).

In the present design, three factors are confounded in generating this effect. While we have not extricated them, it seems well to mention them. The first factor is the autonomous decline noted previously in the control groups. In terms of absolute magnitude, this would seem to account for little of the total effect.

A second factor is forgetting. Were a respondent to judge with the leadership of a culture-bearing confederate, and then to judge in isolation for several sessions, he would in these later sessions be pooling his immediate visual impression of the stimulus with the remembered reports of the other in a previous session. As these memories dimmed, their weighting could be expected to diminish, and the relative weighting of own observation could be expected to increase. No doubt in the present situation forgetting of indoctrination provides some of the decline within the judgment series of a single respondent.

A third possible source of this within-respondents decline in loyalty to the arbitrary culture is the effect of the new group member upon his elders. Since judging was always from eldest to youngest on any given stimulus presentation, this would have to involve an influence carrying over from one turn to the next. We feel sure that some such influence was present, but its demonstration would require a separate experiment placing confederates in the final answering position.

DISCUSSION

In presenting these materials to our colleagues, we have found them most fascinated by the fact of the cultural transmission surviving total replacement of specific individuals, and by the fact of naive respondents becoming unwitting conspirators in perpetuating a cultural fraud. This demonstration was of course our main intent, and this outcome was inevitable if the conformity research upon which we built had any transfer validity. But

while these results are highly significant statistically, they fall short of our expectations. The inculcated arbitrary norms turn out to be eroded by innovation at a much more rapid rate than we had expected. We have hardly provided a laboratory paradigm for the examples of tenacious adherence to incredible superstition with which anthropologists and students of comparative religion provide us.

Of course, a relative weakness in strength is to be expected in a laboratory example. Mature individuals are undergoing transient indoctrination by unknown age-mates, as opposed to the natural setting of the teaching of the very young by tribal adults over a period of years. Fewer confederates are involved than in the tribal situation, although the effect of number upon degree of conformity is unclear in the laboratory. More important, there are no sanctions rewarding conformity nor punishing innovation. Furthermore, it is even possible that the subject of judgment in our experiment is less ambiguous than the subject material of religious belief, for example. There is, after all, a fairly clear consensus among our naive solitary observers. In our situation, the spontaneous deviations from orthodoxy were all in the same direction, whereas for many arbitrary items of culture the deviations may lack such concerted direction.

Nonetheless, the outcome may well warn us against the assumption that a purely arbitrary cultural norm could be perpetuated indefinitely without other sources of support. Even if people weigh the opinions of their elders many times that of their own direct observations, the collective effect of their own observations probably soon erodes a *functionless* arbitrary belief. Where we observed tenacious bizarre cultural belief we must look to more than mere tradition, suggestibility, or conformity, to explain its retention. Latent functions (e.g., Merton, 1949) at the personal or societal level must be present to counteract the pressures from continuous spontaneous innovation in a more natural direction. For example, Moore (1957) and Aubert (1959) have recently suggested such latent functions to superstitious magical lotteries used in the selection of hunting and fishing sites. As our understanding of the requisites of individual and social life further increases, we may expect to discover such latent functions for many if not all of those "meaningless" superstitions which stubbornly persist.

SUMMARY

In an autokinetic judgment situation in which solitary judgments of movement averaged 3.8 inches, confederates were used to establish arbitrary "cultural norms" of 15.5 inches. The transmission of this norm was studied as one at a time confederates and old members were removed from the group while new members were added. Significant remnants of the culture persisted for four or five generations beyond the last confederate. Gradually in each of the 12 experimental groups the arbitrary norm decayed and the group judgments drifted away from it back to the natural norm found in the control groups. The size of group had no clear effect upon the endurance of the norm established, although a group beginning with one confederate and two naive respondents was markedly weaker than the others and transmitted no culture beyond one generation.

Major effects are interpreted in terms of a judgment process in which each respondent weighs the reports of others and his own observations, achieving a compromise between them. Even with the respondents giving the reports of others from 2.5 to 4 times the weight of their own observations, slight "innovations" result which rapidly cumulate to erode the original arbitrary cultural norm. Details of the process and analogues to the cultural transmission of bizarre beliefs are discussed.

REFERENCES

ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie Press, 1951. Pp. 177–190.

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

AUBERT, V. Chance in social affairs. *Inquiry*, 1959, **2**, 1–24.

GERARD, R. W., KLUCKHOHN, C., & RAPOPORT, A. Biological and cultural evolution: Some analogies and explorations. *Behav. Sci.*, 1956, **1**, 6–34.

GOLDBERG, S. C. Three situational determinants of

conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HAGGARD, E. A., & ROSE, E. Some effects of mental set and active participation in the conditioning of the autokinetic phenomenon. *J. exp. Psychol.*, 1944, **34**, 45–59.

KIDD, J. S. Social influence phenomena in a task-oriented group. *J. abnorm. soc. Psychol.*, 1958, **56**, 13–17.

MERTON, R. K. *Social theory and social structure*. Glencoe, Ill.: Free Press, 1949.

MOORE, O. K. Divination, a new perspective. *Amer. Anthropologist*, 1957, **59**, 72.

ROSE, E., & FELTON, W. Experimental histories of culture. *Amer. sociol. Rev.*, 1955, **20**, 383–392.

SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.

SUMNER, W. G. *Folkways*. Boston: Ginn, 1906.

TARDE, G. *The laws of imitation*. (Paris ed., 1890) New York: Holt, 1903.

(Received June 7, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 659-665

DIFFERENCES IN DEVELOPMENTAL EXPERIENCE, DEFENSIVENESS, AND PERSONALITY ORGANIZATION BETWEEN TWO CLASSES OF PROBLEM DRINKERS

RICHARD R. RUDIE[1] AND LAURENCE S. McGAUGHRAN[2]
Moose Lake State Hospital, Minnesota *University of Houston*

A number of responsible people have spent much time and effort in the last decade or so attempting to convey the notion that alcoholism is a disease. Similar attempts have been made over a longer period of time to educate the public about the nature of mental illness. In both cases, the benefit has been a shift in public opinion toward regarding alcoholism and mental illness as preventable and treatable, rather than simply deplorable.

It is possible, however, that the effects of this indoctrination are not entirely salutary. If alcoholism really is a disease, then the direction to be taken in theorizing and experimenting in this area would be to search for the unique set of events—cause, symptoms, course, and sequelae—which, when found, would define it. Most of the research in alcoholism has followed this line of thought. Whether the cause has been sought from genetic, biochemical, dietary, or psychiatric sources, no disease entity has been isolated, and about the only "facts" that have been established concern survey information, symptomatic relief, or fixed chronological stages of drinking behavior. Actually, in patient management, the medical profession has frequently found it expedient to fall back upon the quasi-religious controls of the Alcoholics Anonymous program as its major resource in later stages of treatment.

An alternative to the conception of alcoholism as a disease entity is to consider problem drinking as a consequence of a number of complex patterns of learned needs, attitudes, habits, etc. acquired under varying causal conditions. This general assumption underlies this investigation: Alcoholism is not the name of a specific disease, the nature of which has yet to be fully discovered and understood; rather, it is a loose, inexact sociological or legalistic term used to cover a number of quite different cause-effect sequences.

In a clinical paper written some 20 years ago, Knight (1937) proposed that a clear distinction be made between two classes of problem drinkers—"essential" and "reactive" alcoholics. Knight's clinical description of the essential alcoholic is summarized as follows: He is emotionally dependent, egocentric, pleasure seeking, irresponsible, insincere, and unreliable; he is unable to establish or to maintain long term goals, or intimate, lasting relations with others. Knight's description of the reactive alcoholic is less specifically drawn; but the implication is that the reactive alcoholic, to one degree or another, possesses traits opposite to those listed for the essential group, and in addition is strongly compulsive. Knight assumes that these more generalized differences in behavioral characteristics result from differences in psychosexual development in the early, formative years. But he also characterizes the differences between these two groups in terms of more concrete behavioral implications of the typological distinction. Thus, he refers to likely differences between the groups in terms of educational and occupational experiences, kinds of drinking and eating habits, time of onset of drinking, prognosis, etc.

Although Knight (1937) did not specifically extend his clinical distinction to the following additional points, the investigators judged that they were not inconsistent with his descriptive distinctions. It was inferred that the reactive alcoholic has assimilated cultural values (predominantly middle class) through the process of identification and shows a reasonable degree of the behavior controlling

[1] Formerly at the University of Houston.

[2] Employed part-time at the Houston Veterans Administration Hospital, where part of this study was conducted.

The investigators express their appreciation to N. J. Bradley, Superintendent, H. Davis, Chief Psychologist, M. E. Brandes, Alcoholics Anonymous Counselor, and other counselors and staff members of the Willmar State Hospital, Willmar, Minnesota, for their cooperation and many helpful suggestions during this study.

effects of anxiety and guilt. Contrarily, it was assumed that the essential alcoholic shows a general culture-combativeness, both in his family background and in his own personal history, and is relatively free of the controlling effects of anxiety and guilt.

GENERAL HYPOTHESES

The purpose of this study was to attempt to provide a broader empirical base (or construct validity) for this clinically derived distinction between essential and reactive alcoholics. With the descriptive and inferential material cited as a rough guide, the investigators entered into the choice of instruments and the subsequent analysis of the data on the basis of some general hypotheses.[3]

Family background and marital relations. The essential alcoholic's family (orientation and procreation) background is one of strife, conflict, disloyalty, disunity, and disharmony within the family, and of nonconformity with society and its institutional controls. All of this is less true of the family background of the reactive alcoholics.

General personality characteristics. The essential alcoholic's general personality characteristics are those frequently subsumed under the terms "psychopathic" or "sociopathic" adjustment. This implies more specific traits such as "hedonism" (direct, immediate pleasure seeking), avoidance of social responsibility, and unmonitored expression of feeling or emotion.

Defensive behavior. As contrasted with the essential group, reactive alcoholics employ a more pervasive defensive system. The reactive group's defensive system is relatively complex, both in terms of range of separate defensive tactics for different areas of conflict or threat, and of a relatively greater tendency toward the more exacting tactic—intellectualization.

METHOD

Subjects

Subjects were drawn from a group of male patients undergoing treatment for alcoholism in a large state institution. A total of 34 patients served in preliminary studies to develop the criterion instrument to be described below. Subjects in the main study were drawn from a random sample of 188 male patients to whom the criterion instrument had been routinely administered.

The selection requirements were as follows: AGCT percentile score above the 43rd percentile and scores at least one sigma above or below the mean on the criterion measure (mean of 19.17, *SD* of 5.89). The final group of patients meeting these requirements numbered 36; this was the *N* used for most of the data analysis. At a later time in the study, 22 more subjects were added to facilitate the statistical analysis of data from one instrument—a life history questionnaire. These additional subjects were selected from the original larger group of 188 by resetting the cutting scores on the criterion measure to one-half a sigma above or below the mean.

The correlation values ($N = 188$) between scores on the criterion instrument and age and education did not differ significantly from zero. In the smaller sample ($N = 36$), there were no significant differences between the essential ($N = 20$) and reactive alcoholic ($N = 16$) groups on the variables of age, education, intelligence, or socioeconomic status of the family of orientation. However, in the enlarged sample ($N = 58$), the reactive alcoholic group ($N = 27$), was significantly ($p = .02$) higher than the essential alcoholic group ($N = 31$) in educational level.

Criterion Instrument

Knight's (1937) description of the specific behavioral and situational distinctions between the two types was used to construct an interview form of questionnaire to select the experimental groups. Questions and scoring criteria were developed gradually in two preliminary studies. In final form, the structured inquiry consisted of 14 filler items and 55 scorable items concerning past or present experiences and behavioral tendencies in eight areas (e.g., age at onset of drinking, drinking habits, types of friendship relations, history of gastrointestinal symptomatology).[4] For each subject, a criterion "score" was obtained by summing the number of interview items for which the response was judged to reflect characteristics of essential alcoholism as described by Knight.

[3] Two additional hypotheses set forth in this study concerned group differences in patterns of acquired needs and in traits presumably associated with differences in experience during early psychosexual development. Because of the uncertain validities of the instruments used to test these hypotheses, the completely negative results that were obtained in this study cannot be clearly interpreted. Consequently, this part of the study will not be covered in this report.

[4] The following supplementary materials have been deposited with the American Documentation Institute: a 69-item questionnaire for the selection of essential and reactive alcoholic patients, and a manual for scoring defensiveness with sentence completion data. Order Document No. 6768 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

About a year after the original criterion instrument data were obtained, the questionnaire form was mailed to a stratified subsample of the original 188 subjects with the request to respond again to each of the questions. Nineteen of these 69 subjects returned the completed form; these records were rescored blindly. A rough check on the consistency of the questionnaire data was made by correlating the respective pairs of scores for these subjects; the obtained product-moment value was .66 ($p = .01$). Considering the rather marked differences in the response situations (i.e., responses obtained by interview vs. by mail), this value is regarded as indicative of a satisfactory degree of stability in the measurement of the criterion variable.

Other Instruments and Techniques

The Rorschach test. Rorschach records were judged for type of defensive behavior according to a system proposed by Schafer (1954) and developed as a research technique by Weiss (1956) and Selzer (1956). In this system, the subject's use of form, the content of his responses, and his behavior and attitudes expressed during the test situation, are analyzed to distinguish the most pervasive defensive style among four major modes of defensive operations—repression, intellectualization, projection, and reaction formation.[5] The relatively high interjudge agreement concerning these ratings of defenses from Rorschach data has been previously demonstrated (Eskenazi, 1957; Weiss, 1956). In the present study, one of the raters[6] with previous experience made the judgments of predominant defensive style.

The sentence completion task. Constructed by Wiener, Carpenter, and Carpenter (1956) to distinguish between "repressive" and "sensitizing" defenses as evidenced in verbal behavior, a sentence completion task was used, consisting of 20 items, 5 of which are fillers. The remaining 15 items are designed to elicit five responses in each of three potential conflict areas—sex, hostility, and self-adequacy.

For the purposes of this study, the present investigators developed a new system of criteria for categorizing defensiveness. In this system, responses to the sentence stems were classified under two main headings, "defense" and "no defense." The first major class, defense, subsumes three separate subclasses—intellectualization, repression, and projection.[7] Under the second major class, no defense, are two subclasses—conflict and no conflict. Thus, there are, in all, five separate scoring categories; for each of these, a number of scoring criteria were developed and built into a scoring manual.[8]

In the development of this system of criteria, all of the responses for each item were assembled, and the scoring criteria for each item were established separately, with information identifying the subjects masked out. This procedure was carried out item by item until all responses had been surveyed. Each response, again item by item, was then reconsidered and formally categorized as meeting one of the developed criteria. About 2 months after the formal scoring was completed, a total of 144 responses was rescored. Of these, 134 (or 93%) were judged as meeting the same criterion applied to them in the initial scoring.

The Minnesota-Briggs History Questionnaire. A history questionnaire developed by Briggs (1955) was used to obtain routine case history material. Briggs estimated that this procedure provides about three times as much reliable information as the usual case history interview. Among the areas sampled are family background and relations, home atmosphere, educational and occupational history, dating and sexual history, trouble with authority, and various personality characteristics. Briggs used eight clinical judges to cluster these items into a number of diagnostic, behavioral, and experiential categories. Of these, the following nine clusters were chosen for use here: conflict with parents, unstable-unrewarding home, broken parental home, sibling conflict, psychopathic adjustment, achievement attitudes, educational accomplishment, occupational accomplishment, and parental socioeconomic status. The last category was used as a potential control variable.

A number of relatively minor alterations of the questionnaire were introduced by the present investigators to construct a research instrument meeting the requirements of this study (e.g., revising the items for direct response by the patient). In final form, the revised questionnaire contained 171 items, with the number of alternative responses ranging from three to eight. Five additional item clusters were developed to form the following categories: patient's socioeconomic status, marital difficulty, marital irresponsibility, patient's unmonitored expression of feeling-emotion, and patient's hedonistic orientation.

In the development of a scoring system to yield a total score for each subject for each cluster, separate item-alternative values were checked (and, if necessary, altered) in terms of expected directionality, and then transposed to standard values on a zero-to-nine point scale to equalize the effect of varying numbers of alternatives among the items. For each subject, standard values for each item within a cluster were then summed to obtain the total score for each cluster.

RESULTS

Defensive Behavior

Rorschach data. The distribution of judgments of predominant defensive style based

[5] See Footnote 4.

[5] The type of defensive behavior termed "reaction formation" here has been called "denial" in the previous studies cited. To the present investigators, the term "reaction formation" more accurately conveys the behavioral characteristics associated with this form of defensive maneuver.

[6] The investigators are grateful to Bertram A. Weiss for carrying out these judgments.

[7] In the opinion of the investigators, the pervasiveness of the defense of reaction formation cannot be measured by summing separate responses in a sentence completion task. An essential characteristic of a reaction formation defensive system is a relatively high degree of consistency across a broad range of behavioral transactions. This does not seem to be necessarily as true for other forms of defensiveness.

TABLE 1
DISTRIBUTION OF JUDGMENTS OF PREDOMINANT DEFENSIVE SYSTEM BASED UPON AN ANALYSIS OF RORSCHACH PROTOCOLS

Group	*N*	Mechanism of Defense				Total
		Intellectualization	Reaction Formation	Repression	Projection	
Essential Alcoholics	20	—	—	6	14	20
Reactive Alcoholics	16	1	6	1	8	16

upon an analysis of Rorschach protocols is shown in Table 1. Group differences in the relative *strength* of each of the separate defensive tactics cannot be tested at present because a satisfactory system of estimating these scalar values for each defense has not yet been developed for the analysis of Rorschach protocols. Although an over-all difference between the groups in distribution of *predominant* defensive style seems apparent, it is not possible to assess the appropriate level of confidence for this over-all difference because of low expected frequencies in the distribution. The significance of the group difference in choice of reaction formation as the predominant defense can be estimated ($p = .004$, Fisher's exact tables) and, of course, if the cell frequenceis for intellectualization and reaction formation were combined as representative of more complex forms of defensiveness, the difference would be slightly greater. The higher incidence of repression for the essential alcoholics does not reach significance ($p = .115$, Fisher's exact tables).

Sentence completion data. During the development of scoring criteria for defensive tactics shown in this task, one item of the five in the conflict area of self-adequacy was omitted as not fitting into the scoring schema. Responses to each of the remaining 14 items for each subject were classified as fitting one of the five categories of defensive behavior described above. The results of an analysis of group differences for these data by means of the *t* test are shown in Table 2.

The results clearly indicate that the reactive alcoholics are significantly (.001) more defensive than the essential alcoholics on total defensiveness as measured by the summation of the defense totals of intellectualization, repression, and projection. The reactive alcoholics were higher on each of the separate defense totals, but only significantly (.01) so for intellectualization. Under "no defense," the essential alcoholics were significantly higher than the reactive alcoholics on both "total no conflict" (.05), and "total conflict" (.001). Possibly, it should be observed that these tests of significance are not completely independent of each other since the "defense" and "no defense" total scores are reciprocally related.

Essentially the same results were obtained when responses for each of the three conflict areas were analyzed separately.

The question of relative simplicity-complexity of defensive systems, in terms of differences in range of available tactics between the two groups, was approached from two angles. First, within each of the two groups, the separate defense scores (intellectualization, repression, and projection) were correlated against each other. Only one of the six coefficients was significant; this was between intellectualizing and repressing within the essential alcoholic group ($r = -.58$; $p = .01$). That is, within this group, the over-all tendency to use intellectualizing as a defense is negatively related to the over-all use of repression.

The second form of analysis was carried out by deriving a score for each subject based upon the proportion of the score for the most frequently used defense to the summed score for total defensiveness. Group differences in the distribution of these proportions were then evaluated by computing a chi square. This value was 5.67 ($p = .01$), the essential alcoholic group showing the greater frequency of high proportions. There was some likelihood of a confounding factor in this result because the essential group had lower over-all frequencies in defensive responses and, thus, might be expected to show higher proportions. This possibility was investigated by computing a rank order correlation for all subjects between the proportion value discussed above and the summed score for total defensiveness. The resulting coefficient value (.16) is not significant.

Thus, from two sources, there is the sugges-

TABLE 2

COMPARISON OF DEFENSIVE OPERATIONS BETWEEN THE ESSENTIAL AND REACTIVE ALCOHOLICS ON THE SENTENCE COMPLETION TASK

Scoring Category	Essential Group (N = 20)		Reactive Group (N = 16)		t	p
	M	SD	M	SD		
Defense						
Intellectualization Total	3.00	1.71	4.62	1.58	2.89	.01
Repression Total	2.85	1.80	3.94	1.85	1.72	—
Projection Total	1.60	1.46	1.94	1.08	.78	—
Defense Total	7.45	2.27	10.50	1.70	4.48	.001
No Defense						
No Conflict Total	3.00	1.30	2.00	1.17	2.35	.05
Conflict Total	3.55	1.83	1.50	.94	4.24	.001
No Defense Total	6.55	2.27	3.50	1.70	4.48	.001

TABLE 3

SIGNIFICANCE OF DIFFERENCES BETWEEN ESSENTIAL AND REACTIVE ALCOHOLICS ON ITEM CLUSTER CONVERTED SCORES FROM THE MINNESOTA-BRIGGS HISTORY QUESTIONNAIRE

Title of Cluster	Number of Items	Essential Group Mean (N = 31)	Reactive Group Mean (N = 27)	Chi Square	p
Family of Orientation					
Conflict with parents	22	83.80	77.21	6.08	.02
Unstable-unrewarding home	18	58.88	51.41	9.43	.01
Broken parental home	9	30.29	28.72	.39	—
Sibling conflict	6	24.53	21.55	4.52	.05
Family of Procreation[a]					
Marital difficulty	22	64.13	57.90	14.70	.001
Marital irresponsibility	4	31.68	26.09	7.04	.01
Socioeconomic level	3	14.12	17.12	8.71	.01
Personality Characteristics					
Unmonitored feeling-emotion	14	64.23	58.76	5.81	.02
Hedonism	13	57.44	49.15	8.98	.01
Psychopathic adjustment	23	87.79	73.94	15.59	.001
Unfavorable achievement attitudes	17	71.93	62.47	15.59	.001
Educational accomplishment	9	42.16	47.22	6.42	.02
Occupational accomplishment	5	25.84	27.84	4.97	.05

[a] For items concerning family of procreation, N = 27 in each group.

tion that the essential alcoholic group employs a relatively simple, generalized defensiveness, and the reactive alcoholic group a relatively complex, more specialized system of defenses.

History Questionnaire Data

Data concerning the subject's family background, marital relations, and general personality characteristics were obtained from the 13 item clusters of the history questionnaire. The results of the analysis of these data by chi squares are shown in Table 3. With all 13 clusters, the group differences are in the expected direction, and in 12 of these comparisons, the differences are significant at p levels of .05 or beyond. These results are not completely independent because there is a certain amount of item overlap among some of the clusters; however, this overlap is not extensive.

DISCUSSION

Although previous experimental and clinical findings provided no basis for specifically anticipating a higher incidence of the defense of reaction formation among the reactive alcoholics, this finding is not too surprising, in view of the relative complexity of this defensive technique.

One finding involving the Rorschach data, however, was clearly not anticipated. This was the lack of comparability in the results of the Rorschach and sentence completion data analyses of defensive behavior. There is evidence of reasonable interjudge agreement for each of the two general systems of judgment

of defensive behavior. Further, each of these two systems has led to successful predictions of behavior in other studies. Thus, some network of meaning exists for the concepts of defensiveness tied into each set of operations. Assuming that the two techniques for inferring kinds of defensive behavior are reasonably reliable, one is led to the conclusion that the Rorschach and sentence completion situations represent actually different sets of response requirements.

The most obvious difference in the two testing situations concerns the degree of structure. It follows that differences in degree of structure in the testing situations would be associated with differences in degrees of awareness in the subjects of the implications of their test behavior to the examiner in the two situations. One of the clearest differences in the Rorschach and sentence completion records in this study was that of a high incidence of judgments of projection on the Rorschach and a relatively low frequency of responses reflecting this defense in the completion task. Technically, it was quite possible for the subjects to complete each of the items in the latter task by employing a projective defense. The fact is, they usually did not. Similarly, it was quite possible to reply to the sex-conflict items by openly acknowledging a conflict in this area. Again, this type of response simply was not given.

Under ordinary social conditions, the use of projection as a defense is a luxury few can afford. Its cost, usually, is social punishment and a consequent loss of "face" if detected, or some degree of loss of reality, if not. It is one of the earliest temptations for a young child and, frequently, one of the first response tendencies to evoke strong negative sanction. In Cameron and Magaret's (1951) terms, parental disapproval of the child for falsely imputing blame to others would frequently be one of the earliest conditions for the development in the child of "muting of communicative speech" and, later, "behavioral duplicity."

The instructions in the completion task are "to express your real feelings." It appears likely that, to varying degrees, subjects reinterpret this instruction, when threatened, to the personal meaning of "suppress your real feelings, if you can do so and still appear to comply." Thus, the avoidance of too-obvious projection, the avoidance of acknowledgment of sexual conflict, and the apparent but not actual compliance with test instructions to express "real feelings" are all tactics of "behavioral duplicity" that are relatively "successful." They are segments of well-learned behavior sequences which permit the individual to "play his roles"—to escape threat while still maintaining self-integrity in the eyes of the observer.

But, as has been frequently observed, the "rules of the game," or the cultural expectancies, are much less clear in the less structured Rorschach situation. If this interpersonal situation is sufficiently threatening to the individual, he has fewer guide lines to lead him toward the safer defensive tactics and he may, in desperation, resort to the more dangerous projection—to "attack" or impute blame to others, or to acknowledge openly his fear of "attack" from others.

One of the tentative findings of this study was that the specificity-generality of a defensive system may be a function of the complexity-simplicity of organization of the individual personality. Here, the possibility is also suggested that choice of defense may vary with the individual's degree of awareness of the implications of his behavior. There may be other determinants of defensive choice (e.g., degree of threat, amount of skill and experience the specific defensive technique might require to be successfully used, or type of conflict) that could be introduced systematically in future research.

Most of the discussion has dealt with results obtained in this study that were essentially unanticipated. A good proportion of the results were anticipated in terms of Knight's (1937) clinical observations concerning the distinctiveness of essential and reactive alcoholism. The classes of expectancies that were, in good part, confirmed concerned group differences in defensive behavior, developmental experiences, and a number of aspects of personality organization. Thus, Knight's distinction between essential and reactive alcoholics appears to be an essentially valid and useful one, which can serve to reduce variance in future clinical research in much the same way that such other clinical distinctions as process-reac-

tive, good and poor prognosis in schizophrenia, and satellizing-nonsatellizing in personality development have already done.

SUMMARY

On the basis of clinical experience, Knight (1937) has proposed a distinction between two classes of alcoholism—"essential" and "reactive." The purpose of this study was to attempt to provide a broader empirical base for this distinction in terms of differences in developmental experiences, defensiveness, and personality organization.

A pretested selector questionnaire developed to establish the criterion groups contains 55 scorable items in eight behavioral areas emphasized by Knight (e.g., drinking habits, types of personal and institutional dependency, history of gastrointestinal symptomatology). Subjects were males hospitalized for alcoholism. There were no significant group differences in age, intelligence, or parental socioeconomic status.

As expected, the reactive alcoholics showed a higher degree of over-all defensiveness, their defense system was more complex (less generalized), and they placed greater reliance upon the more exacting tactic—intellectualization. Also as expected, the essential alcoholics "scored" significantly higher on psychopathic adjustment, conflict with parents, unstable-unrewarding home, sibling conflict, hedonism, unmonitored feeling-emotion, marital difficulty, and marital irresponsibility. The reactive alcoholics "scored" significantly higher on over-all achievement, educational accomplishment, occupational accomplishment, and socioeconomic level of the conjugal family.

Knight's distinction between these classes of alcoholism appears to be an essentially useful and valid one that can serve to reduce variance in future research with alcoholics.

REFERENCES

BRIGGS, P. F. Preliminary validation of a standard personal history for psychiatric diagnosis. Unpublished doctoral dissertation, University of Minnesota, 1955.

CAMERON, N., & MAGARET, ANN. *Behavior pathology.* New York: Houghton Mifflin, 1951.

ESKENAZI, A. Personality development of pre-adolescent boys as a function of mothers' defenses and parental attitudes. Unpublished doctoral dissertation, University of Houston, 1957.

KNIGHT, R. P. The dynamics and treatment of chronic alcohol addiction. *Bull. Menninger Clin.*, 1937, **1**, 233–250.

SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

SELZER, S. Relationships between developmental experiences and choice of defensive behavior: Study II. Females. Unpublished doctoral dissertation, University of Houston, 1956.

WEISS, B. A. Relationships between developmental experiences and choice of defensive behavior: Study I. Males. Unpublished doctoral dissertation, University of Houston, 1956.

WIENER, M., CARPENTER, B., & CARPENTER, JANETH T. Determination of defense mechanisms for conflict areas from verbal material. *J. consult. Psychol.*, 1956, **20**, 215–219.

(Received June 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 666-673

THE ASSESSMENT OF ANXIETY IN PSYCHOLOGICAL EXPERIMENTATION:

A THEORETICAL REFORMULATION AND TEST[1]

GEORGE C. ROSENWALD
University of Michigan

THE effects of anxiety on performance have been subjected to fairly intensive experimental study in past years (Child, 1954). A good deal of this research has shed light on the type of performance that is disrupted by anxiety. For instance, it has been shown that anxiety interferes with the accomplishment of complex tasks but facilitates the performance of simple tasks (Farber & Spence, 1953). Considerably less attention has been given to a theoretical analysis of anxiety itself, the conditions of its arousal, and its defining properties. Experimenters do not always agree with each other or with theoreticians and clinicians as to the exact meaning of anxiety as a conceptual term. The first part of the present paper is a review of the psychoanalytic theory of anxiety. This framework is selected because many authors have operated within it. Following the presentation of the theoretical outline, some shortcomings of conventional experimental approaches to anxiety will be pointed out. Finally, an experiment will be reported, the procedures of which approximate the theoretical specifications more closely.

The first of the two conventional procedures to be discussed segregates the subjects with respect to anxiety level on the basis of a questionnaire to which the subject responds with an introspective report of his characteristic anxiety level and stress experiences (Montague, 1953; Taylor & Spence, 1952). The second procedure produces a stress in the subject through fictitious reports of his inadequate performance on tasks which are presented as tests of intelligence or as other measures of adequacy (Mandler & Sarason, 1952; Sarason, Mandler, & Craighill, 1952). Sometimes these two procedures are employed jointly. They are subject to limitations which will become clear after the theoretical outline is presented.

The psychoanalytic theory of anxiety (Freud, 1936) begins with the observation that for some people the intensification of a particular drive constitutes a psychic danger. The nature and origin of this danger are to be sought in the individual histories of such persons. Unconsciously fantasied physical damage, helplessness, or loss of love are frequently encountered examples of such danger. In anticipation of this danger, such people characteristically distort the recognition and expression of the drive in question by means of defense mechanisms. These mechanisms avert the expected danger to the extent that they block or distort the drive successfully. Anxiety is defined as the psychological mechanism whereby the current intensification of a dangerous drive results in the elicitation of defenses. The term "signal anxiety" refers to this function specifically. Anxiety is therefore a theoretical construct which is anchored on the antecedent side to the intensification of a dangerous drive and on the consequent side to the rise of defensive behavior.[2] It should be noted that this strict definition of anxiety

[1] An abbreviated version of this paper was prepared for presentation at the International Congress of Psychology in Bonn, Germany, 1960. It is based on a dissertation presented to the faculty of the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy.

The author is grateful to Irving L. Janis, William Kessen, and Roy Schafer for their advice and assistance throughout the conduct of this study, which was carried out while the author was a Public Health Trainee in Clinical Psychology.

Thanks go to Cynthia Dember for her assistance in the rating of story completions as well as to Carl Garvin, principal, and the students of the East Haven, Connecticut, High School for their helpful cooperation.

[2] Anxiety as a mechanism should not be confused with anxiety as a character trait. To say "John is anxious about aggression," does not necessarily mean that John is aggressively stimulated at the present time and is counteracting this stimulation with defenses. Rather, it means that John would become defensive if he were aggressively stimulated. This is another way of saying that aggression is a dangerous drive for John or that John has conflicts about aggression.

makes no phenomenological references to feelings of distress, heart palpitations, etc.

Defenses are therefore a central concept in the definition of anxiety. They are recognized when the individual behaves in ways which interfere with the experience or expression of his needs. For example, when it is observed that a person responds to extreme aggressive provocation with inexhaustible forbearance and forgiveness for the offender, it may be said that he is defending himself against aggression. In clinical practice and experimental studies (Eriksen, 1954) it has been found that people who are defensive characteristically distort their perception of stimuli which are associated with the arousal of dangerous drives. Moreover, it is found that the greater the danger, the greater is the distortion (Schafer, 1954).

It is well known that anxiety is in certain instances further characterized by conscious experiences of distress and by physiological stress symptoms. However, these additional conditions are of secondary importance in the definition of anxiety and may be completely lacking. Thus the term anxiety may be applied regardless of whether subjective experiential factors and physiological stress symptoms play a minimal role or assume spectacular importance, as they do, for instance, in anxiety attacks. Only its traceable origin in past experience, which has rendered the drive dangerous, and the automatic elicitation of defenses are central to the definition of anxiety. In fact, the concomitance of anxiety, whatever its experiential salience, and of defensiveness is so constant that it is difficult to distinguish conceptually or experimentally between the behavioral consequences of the one and the other. Under what conditions do the secondary experiential and physiological factors appear at all? The theory provides an explanation in terms of the effectiveness of the defense. If the defense mechanism does not sufficiently block or distort the dangerous drive, and the drive grows in intensity, then the experiential and physiological aspects of anxiety will gradually become more conspicuous. In other words, in the case of a person with ineffective defenses, these aspects will be added to signal anxiety when the drive is intensified. They subside as soon as the drive is sufficiently distorted.

It is an important part of the theory that anxiety operates automatically in many cases. That is to say, the elicitation of defenses does not depend on the experiential and physiological aspects of anxiety. This is especially characteristic of functioning normal adults. In fact, people who are frequently subject to anxiety *experiences* are often in need of treatment. Anxiety is not an ever-present condition or mood in well functioning individuals. Even in patients, the intensity and effects of anxiety and defensiveness vary with the intensity of the drive that gives rise to them. Only an acute intensification of dangerous drives will result in anxiety—experiential or otherwise. For instance, a person for whom aggressive stimulation is dangerous will try not to expose himself to social situations in which he may be aggressively stimulated. Thus anxiety and defense arousing situations may only rarely occur in his life.

The view outlined here reserves the concepts of "anxiety" and "defense" for situations in which the danger has an internal source, specifically a drive, and the term "fear" for those in which the threat originates in the external world. Although borderline instances and cases of overlapping may be cited, the paradigm itself is clear in this respect.

In the light of this theory, both the questionnaire and the stress procedures mentioned initially have serious shortcomings. As for the questionnaire, its outstanding limitation is its reliance on the subject's introspective report of anxiety. By theory, subjects who rate themselves high on anxiety by reporting experiences of psychic distress and physiological stress are not only anxious in the primary sense of the word—they not only become defensive in the presence of the drive—but in addition, they utilize ineffective defenses which do not distort the drive sufficiently. Furthermore, they may not be successful in avoiding situations in which they are likely to be dangerously stimulated. While these individuals represent a challenging clinical problem, it is incorrect to designate others, who do not report experiences of anxiety, as not anxious in the strict sense. The only admissible distinction between subjects who report and subjects who do not report anxiety experiences is on the basis of the effectiveness of their defenses, not on the

basis of whether or not defenses are being elicited. Subjects who do not report anxiety experiences are either not anxious about the drive, or they avoid anxiety arousing situations; or, if they do not avoid them, then anxiety never becomes more than a signal because their defenses are effective. It should be noted that people who resolve underlying conflict with relatively flexible defenses are not detectable by means of a questionnaire.

As for the stress procedure, it seems to approximate the specifications of fear arousal more than those of anxiety arousal. True, the prospects of failing an intelligence test or of showing up poorly on a measure of masculinity are dangerous both internally and externally; such failures would involve a compromise in the eyes of others as well as a drop in self-esteem. Yet it seems preferable to investigate these effects separately rather than under confounded conditions. If it were possible to bring about the intensification of a dangerous drive without simultaneously producing an immediate external threat, the conditions of anxiety arousal specified in the theory would seem to be most nearly met.

In view of these considerations, the following method was proposed. Aggression was selected as a drive about which people are anxious to varying degrees.[3] First, the degree of anxiety about aggression was to be determined for each subject by noting his extremity of defensive distortion in response to an aggressive assessment stimulus. Then, independently of this determination, subjects were to be exposed to an acute aggressive provocation. It was assumed that anxiety and defensiveness would automatically be aroused again, just as they were in response to the assessment stimulus, and that the effects of anxiety and defensiveness on behavior could then be measured. In other words, the subject's own "built-in" predisposition to become anxious as a result of drive arousal was to be utilized rather than a fictitious external threat. Nor was any use to be made of the subjects' awareness of their anxiety.

An anagram solving task was selected as a measure of the effects of anxiety on performance. Findings indicate that the effect of anxiety upon performance varies with the complexity of the task (Farber & Spence, 1953). A task is considered complex if the correct response is either weak in the initial response hierarchy or is in competition with other equally strong responses, whereas a simple task, like that in classical conditioning, is one in which the correct response is strong in the hierarchy. In the former tasks, anxiety has been found to have an impairing but in the latter a facilitating effect on performance. Anagram solution is a complex task which is known to be vulnerable to the effects of anxiety as measured by questionnaire (Maltzman, Fox, & Morrisett, 1953).

The prediction tested by the present experiment is the following: In the case of subjects who are observed to become defensive in reaction to an aggressive assessment stimulus, an experimental intervention designed to arouse aggression results in impaired performance on a complex problem solving task. Subjects who do not become defensive in response to the assessment stimulus are not so affected by the experimental intervention.

METHOD

The experimental subjects were 62 high school students who were tested individually during two separate sessions. During the first session, a Story Completion test, an Anagrams test, and an abbreviated Intelligence test were administered (Rosenwald, 1958). Thirty-nine of the subjects constituted the experimental group. At the beginning of their second session, they were subjected to an experimental intervention designed to arouse aggression. For the remaining 23 control subjects, this intervention was omitted. Following this, alternate forms of the three tests were administered to all subjects. Differences between the forms of the tests were controlled for by the use of a counterbalanced design.

The Story Completion test consisted of five incomplete stories which the subject was asked to complete. Each contained an aggressive provocation of the main figure. This thematic test was selected as the assessment stimulus because projective devices are traditional for detecting anxiety and defensiveness in the sense defined, and because responses to such tests are ordinarily less subject to influence by socially conditioned attitudes. Such tests may identify individuals who characteristically respond to increases in drive in a defensive manner but are not necessarily aware of their anxiety or of the entailed defensive drive distortion. Because projective test responses reflect anxiety and defensiveness concerning which the subject has little or no control or awareness, they seem particularly suited to the present experimental paradigm.

[3] The choice of aggression is not crucial to the theoretical formulation of this paper. Anxiety about dependency or sexuality could have been utilized as easily. This will be discussed later on.

The Anagrams test consisted of 10 scrambled four- and five-letter words which the subject was asked to convert into meaningful words by rearranging the letters in his mind. Half the anagram solutions were "loaded"—that is, they had aggression related connotations; the other half was neutral. According to the conception of the present study and the findings of past research (Maltzman et al., 1953), the anagrams task should be subject to the effects of anxiety regardless of the content of the anagrams. The rationale of this prediction is entirely in terms of task complexity, and in this respect it differs from studies of perceptual defense. "Loaded" stimuli were included out of interest in the effects which content may perhaps have beyond those postulated. Just as incomplete stories with aggressive themes are more likely to point up a subject's anxiety about aggression than stories with other themes, so the aggressive content of the anagrams may add to the anxiety which a subject already experiences as a result of the experimental drive arousal.

The Information subtest of the Wechsler Adult Intelligence Scale was administered as a measure of intelligence (Wechsler, 1955).

The experimental intervention consisted of two parts. First, the subject was accused of and reprimanded for having discussed the experiment with his friends after having been asked at the end of the first session to refrain from doing so. He was told that this was childish, disobedient, and irresponsible. Second, he was asked to take a mirror tracing test as part of the study. His performance in this was characterized as inadequate, and the experimenter demanded that the subject return after school to continue practicing mirror tracing. The intervention was delivered in a sarcastic, belittling manner, and no replies were tolerated. (Mirror tracing, a measure external to the experiment proper, was chosen for the intervention because of possible changes in task set which might have affected performance on the Anagrams test had *it* been used for this purpose.) All of the subject's verbalizations at this time were recorded by the experimenter. The demand for after school practice was retracted at the end of the experiment, and the overall purpose of the study was explained.

RESULTS

The five story completions obtained in the first session were used to obtain an index of defensiveness. Each completion was ranked with respect to defensiveness in relation to all others obtained in response to the same incomplete story. A completion was judged to be defensive in proportion to the extent to which aggression was *not* expressed, the extent to which the main figure of the story was *not* seen as active in bringing about a favorable solution to the presented conflict, the extent to which he was *not* seen as capable of acknowledging the discrepancy of interest between himself and the instigator of the aggressive provocation, and the extent to which he was *not* seen as blaming the instigator but rather as blaming himself. No attempt was made to categorize the completions according to specific defense mechanisms exemplified by them. Rather, the degree of defensiveness was determined by the magnitude of the distortion. By using an objective scoring manual (Rosenwald, 1958), an alternate judge arrived at reliably similar rankings, with rho coefficients ranging from .71 to .92. The ranks obtained for each of the subject's five stories were summed, yielding his defensiveness score. The subjects were then divided at the median into a high defensive and a low defensive group.

TABLE 1

EXTENDED MEDIAN TEST (Siegel, 1956) AND WILSON'S DISTRIBUTION TEST (Wilson, 1956) FOR THE DIFFERENCE BETWEEN SOLUTION TIMES ON THE FIRST AND SECOND ADMINISTRATION OF THE ANAGRAMS TEST

(High scores indicate relative impairment)

	Experimental		Control	
	High-Defensive	Low-Defensive	High-Defensive	Low-Defensive
Above Overall Median	14	4	6	7
Below	5	16	5	5
χ^2 Total	11.90		3*df*	$p < .001$
χ^2 Defensiveness	6.46		1*df*	$p < .02$
χ^2 Treatment	0.62		1*df*	—
χ^2 Defensiveness × Treatment	4.82		1*df*	$p < .05$

Note.—A nonparametric test was employed to permit the pooling of difference scores obtained from two replications differing in the order of administration of parallel test forms and therefore not comparable in absolute values of solution time.

In comparing the times required for solution on the first administration of the Anagrams test, it was found that high defensive and low defensive subjects performed comparably.

An analysis of performance changes from the first to the second administration of the Anagrams test is shown in Table 1. The breakdown of the chi square into the components attributable to Defensiveness, Treatment, and their interaction was undertaken, using Wilson's (1956) test. It was concluded that the performance changes of the high defensive and low defensive groups were significantly different, the former showing impairment and the latter improvement of performance when compared to the control subjects. Furthermore, these

changes could be attributed to the effects of the experimental intervention since the interaction of Defensiveness and Treatment was found to be significant.

In addition, there is suggestive but statistically unreliable evidence that high defensive subjects required longer solution times for "loaded" than for neutral anagrams on the *first* administration of the test and that this was not so for the low defensive subjects. However, the reported effects of the experimental intervention were not limited to the loaded items. If anything, neutral items seemed to be affected somewhat more noticeably by the intervention than were loaded ones.

The high defensive and low defensive subjects did not differ in their performance on the intelligence scale either before or after the intervention.

DISCUSSION

The above findings may be briefly summarized as follows: High defensive and low defensive subjects do not at the outset differ from each other in regard to their performance on the Anagrams test; they are differentially affected by the experimental intervention, with the high defensive subjects showing impairment and the low defensive subjects improvement as measured by solution times, and the content of the items does not seem to influence their susceptibility to the facilitating or impairing effects of the experimental intervention.

In connection with the first of these three findings, it should be noted that investigators assessing anxiety with questionnaires have found that subjects differ in performance even in the absence of any experimental arousal (Farber & Spence, 1953; Taylor & Spence, 1952). The failure of the present study to corroborate this observation may be taken as indirect evidence that Defensiveness rankings are not interchangeable with scores on an anxiety questionnaire, as indeed they were not intended to be. The present findings can be accounted for without assuming that any of the subjects ever became consciously distressed or experienced physiological stress symptoms, although such a contingency is certainly not ruled out. Since the effectiveness of defenses is theoretically related to the presence or absence of the experiential aspects of anxiety, it is possible that differences in performance which have been found in past studies between subjects high and low on an anxiety questionnaire are due to the degree of defense effectiveness rather than to the degree to which the subject becomes anxious in the theoretically strict sense. The present study is only a first attempt to elucidate experimentally the differences between the behavioral consequences of overt anxiety, introspectively observed by the subject and reported on a questionnaire, and of the disposition to become defensive, measured by projective devices and not necessarily accompanied by conscious experiences of distress.

An inspection of the items of the Taylor (1953) scale suggests that a high score is likely to indicate a more or less permanent drive or affective state. The rationale and findings of the present study, however, explicitly provide for and give positive evidence of an acute rise in anxiety conditional on an acute rise in a dangerous drive. Once again, it appears that the present procedure for controlling anxiety could be applied to people who are not characteristically subject to the experiential accompaniments of anxiety, even though they may be judged anxious and defensive about a particular drive. The method described here might, therefore, facilitate the conceptual disentanglement of the effects of anxiety and defense.

The first and second findings, taken together, suggest strongly that the defensiveness rankings measure what it was hoped they would measure: not the person's usual or typical level of anxiety, but his disposition to become anxious and to defend when dangerously aroused. In accordance with the rationale of this study, it should be possible to replicate the reported findings by raising the level of any other drive in subjects for whom that drive is known to be dangerous.

With regard to the second finding, the question arises as to what psychological mechanisms were responsible for the observed performance changes. It may be argued that a high defensive subject could so distort the meaning of the experimental intervention and of his reaction to it that he might ignore its provocative implications. After all, the intervention is a stimulus associated with drive arousal just as the incomplete stories were.

In that case, should the intervention not fail to affect him? This is certainly possible, but it must be remembered that anxiety always occurs in combination with defensive efforts of varying kinds and degrees of success. At the present state of knowledge, the effects of anxiety and defense on cognitive behavior cannot yet be distinguished either experimentally or even conceptually. In other words, the high defensive subjects' test performance was impaired by anxiety or the defensive distortion of the experimental intervention or both. Parallel with this, it is unclear whether investigators of anxiety who have not been systematically concerned with the effects of defense were justified in attributing their findings only to the effects of anxiety. Perhaps their data should be viewed in part from the standpoint of the defenses which are automatically aroused by the anxiety stimulus.

The separation of the effects of anxiety from those of defense would have more than semantic importance. It is conceivable that anxiety affects performance through the mediation of drive increments and irrelevant responses (Child, 1954), whereas defense may exert its influence through the mediation of cognitive rigidity and the narrowing of attention and associative processes.[4] Whether or not a particular task performance is affected by anxiety and defensiveness would then depend on the effectiveness of an individual's defenses and on the vulnerability of the selected task to one or the other type of mediating mechanism. For instance, if performance of a task is most vulnerable to cognitive rigidity, then a subject who is only moderately defensive and experiences a good deal of distress will not be so severely penalized in his performance as another subject whose defenses are intensely and pervasively constricting and who is, therefore, free from experiences of overt distress and physiological stress. The analysis of intellectual functioning in different diagnostic groups tends to bear this out (Rapaport, Gill, & Schafer, 1945).

Little is known at the present time that would easily account for the improvement of the low defensive subjects. It is possible that unanxious, and therefore undistorted, experience of increased aggression provided them with greater energy, perhaps even with vindictive overtones. ("I'll show him who is immature!") Although the finding is not surprising in the light of clinical experience, there is little if any research support for the assumption that the aggressive drive can at times be utilized in this constructive way. Another possible explanation is that this added drive had merely a distracting effect on the low defensive subjects and that they overcame it by intensifying their efforts to attend only to the task. Whether or not *any* drive arousal will have such a facilitating effect on subjects who experience little anxiety about it is not known. It is quite possible that aggression, because of its nature, is more likely to have such energizing effects than is, say, an arousal of dependency needs. Furthermore, one of the criteria used for categorizing subjects as high defensive or low defensive was whether or not they made reference to active problem solving behavior on the part of the main figure in their story completions. The low defensive subjects' superiority on the Anagrams test may be looked upon as another facet of their greater concern with or investment in problem solving. Just as these subjects involved their story figures in active problem solving attempts, so they became themselves involved in active attempts to restructure the anagram stimuli.

The findings of the experimental groups are interesting also in that they are consistent with economic considerations in psychoanalytic theory. It is assumed that the ego performs its adaptive functions, of which problem solving is one, with energy which is of instinctual origin but has undergone varying degrees of neutralization. The capacity to neutralize instinctual energy is thought to depend on the person's relative freedom from anxiety concerning the instinct in question. That is, defensiveness tends to preclude neutralization. The present findings can be explained as follows. The low defensive group had a greater quantity of neutralized, dischargeable energy available for problem solving and therefore showed an improvement, whereas the high defensive group had to block the aroused aggressive energy and therefore pre-empted it from adaptive utilization.

[4] The effect of rigid defenses and controls on one phase of creative and inventive behavior has been discussed by Schafer (1958).

These economic considerations seem the more applicable because the groups performed comparably in the absence of energic arousal and because the stimulus content was noncontributory to the observed effects.

An incidental observation needs to be mentioned in connection with the second finding. An analysis of the experimental subjects' verbal reactions to the experimental intervention revealed a tendency for high defensive subjects to object to the experimenter's accusation and arbitrariness more frequently than was true of low defensive subjects. Although this finding falls far short of statistical significance, failure to find a significant difference in the opposite direction serves as a reminder that one may not freely predict diffident and meek behavior from the presence of defensiveness in projective test responses. In many everyday situations, such defenses are overlaid with socially conditioned counterdefenses against passivity and timidity. In neurotic patients, considerable therapeutic work may often be required to bring to the surface the submissiveness which hides behind impressive swagger and assertiveness. However, this relationship between fantasy defensiveness and action assertiveness stands in need of further exploration.

The third finding suggests that theories derived from the phenomena of perceptual defense may not provide adequate explanations of the present observations. Apparently, psychological functions like those involved in the solution of anagrams can be subject to wholesale interference which transcends particular areas of tabu content. It is suggested that although anxiety is aroused by drives which have specifiable content, the effects, facilitating or inhibiting, of anxiety and defensiveness on performance may have a wider sphere of influence than the anxiety arousing content. There is no a priori reason to assume that anxiety will affect a given type of cognitive behavior differentially, depending on what drive gave rise to the anxiety. This too is borne out by the analysis of intellectual functions in patient groups (Rapaport et al., 1945).

Finally, it could be argued that the effect of the experimental intervention operated through a decrease in the cooperativeness or incentive of the high defensive subjects. The best reply to this and similar explanations is a question: Why should the experimental treatment have different effects on the cooperativeness of people who have high and low defensiveness ratings? The advantage of combining predispositional and experimental controls is that the number of plausible explanations of the obtained effects can be greatly reduced.

SUMMARY

The psychoanalytic theory of anxiety and defense was reviewed. Following this, weaknesses were pointed out in experimental methods which apply the term "anxiety" to overt experiences of subjective distress accompanied by physiological distress symptoms, or which seek to control the level of anxiety by means of external threats. A new procedure was proposed and tested, consisting in the experimental arousal of a motive known to be psychologically unacceptable or dangerous for the subject. Anxiety is thought of as the automatic accompaniment of such arousals, regardless of its experiential conspicuousness to the subject himself.

A thematic test was utilized to assess defensiveness about aggression. Subsequently, an experimental intervention, designed to arouse aggression, was administered to the subjects, and the effects of this intervention on the performance of a problem solving task were measured. Findings were as follows:

1. Under drive arousal, high defensive subjects showed impairment and low defensive subjects improvement in performance.

2. Subjects did not differ in level of performance before drive arousal.

3. The content of the stimulus material used to measure problem solving ability did not seem to contribute to the vulnerability of performance to the effects of anxiety and defensiveness.

This reformulation of anxiety appears to hold promise for the conceptual and experimental separation of the effects of anxiety and defense.

REFERENCES

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

FREUD, S. *The problem of anxiety*. (Originally published 1926) New York: Norton, 1936.

MALTZMAN, I., FOX, J., & MORRISETT, L., JR. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. I. New York: International Univer. Press, 1945.

ROSENWALD, G. C. The effect of defensiveness on ideational mobility. Unpublished PhD dissertation, Yale University, 1958.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.

SCHAFER, R. Regression in the service of the ego: The relevance of a psychoanalytic concept for personality assessment. In G. Lindzey (Ed.), *The assessment of human motives*. New York: Rinehart, 1958. Pp. 119–148.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

WECHSLER, D. *Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received June 15, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 674-682

DIMENSIONS OF INDIVIDUAL PERFORMANCE IN SMALL GROUPS UNDER TASK AND SOCIAL-EMOTIONAL CONDITIONS[1]

RICHARD D. MANN[2]
Harvard University

WITHIN recent years considerable attention has been directed to the search for a minimum number of dimensions in terms of which to describe a person's overt behavior and informal status in small groups. Perhaps as a natural antidote to the growing number of performance measures, factor analytic techniques have been employed in an effort to demonstrate that a large proportion of the variance of these performance measures may be accounted for by a small number of underlying dimensions. The major advantages of being able to specify such a set of factors include greater economy and efficiency of effort in measuring performance, greater consistency of research methods across studies, and the opportunity to consider the theoretical significance of the factors that are found to be of primary importance.

Carter (1954), in his review of several factor analytic studies of individual performance, concluded that three factors recur with sufficient regularity to constitute at least a preliminary set of dimensions for the field. He labeled these factors: Individual Prominence and Achievement, Aiding Attainment by the Group, and Sociability. Bales (1956), in his discussion of Carter's synthesis, proposed that the factors be labeled: Activity, Task Ability, and Likeability. However, subsequent studies have resulted in less than perfect agreement with the three-factor scheme proposed by Carter and Bales. Borgatta, Cottrell, and Mann (1958), in their factor analysis of 40 personality traits and behavioral descriptions, found five factors, three of which correspond to Carter's proposed set. The remaining two factors they label Manifest Emotionality and Task Interest. On the other hand, Slater (1955) found only two significant factors of performance in small groups engaged in discussing human relations problems. His first factor, Task Prominence, represents a fusion of Carter's first two factors, those which Bales calls Activity and Task Ability; Slater's second factor, Likeability, is in close agreement with Carter's Sociability factor. Thus, the three factors proposed by Carter and Bales appear under some conditions to underestimate and under other conditions to overestimate the number of significant dimensions of performance that are found.

Unfortunately, the research methods used in the various studies have been so divergent that any attempt to make use of previous research to determine the critical conditions under which the nature and number of performance factors vary is bound to be a hazardous enterprise. Among other things, the groups studied by the various investigators have differed in the task assigned to them, the length of previous interaction, size, the use of observer scoring techniques vs. peer ratings, and the nature of the populations from which the samples were drawn. While Carter's (1954) list may be a useful beginning, the question of how stable such a set may be under various conditions clearly deserves more attention.

The purpose of the present research was to determine what effect varying the group problem would have upon the dimensions of individual performance. The group problems were chosen with direct reference to the dichotomy drawn by Bales (1950) and Parsons (1951) between task groups in which adaptive-instrumental goals predominate and social-emotional groups in which integrative-expressive goals predominate. In Maier's (1952) terms, task groups are more concerned with the quality of their solution, whereas social-emotional groups are more concerned with reaching maximum consensus. The main question for this study, then, was whether the underlying dimensions of performance remain constant across task and social-emotional conditions.

[1] This article is based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Michigan.

[2] The author wishes to express his appreciation to Roger W. Heyns for his guidance and encouragement.

METHOD

Subjects

The subjects were 100 male undergraduates at the University of Michigan, drawn from the most recently initiated members of 16 undergraduate social fraternities. With the exception of 14 sophomores and 1 senior, all the subjects were freshmen. The subjects were recruited through their pledge-class presidents and were offered $5.00 for roughly 5 hours of experimental time.

Experimental Groups

Each subject participated in two five-man group discussions, once under each experimental condition, no two subjects appearing together in both task and social-emotional groups. Before the groups were composed, each subject was asked to indicate which of the other 99 subjects he knew. Of the 400 pairs of subjects in the 40 experimental groups, in only 15 did one member say he had a passing acquaintance with the other member; in only three of these cases was the recognition reciprocated. The experimental sessions were conducted in the Group Dynamics Laboratory at the University of Michigan, which is equipped with an observation room with a two-way sound system and a one-way observation screen. Half of the subjects were placed in the Task condition first, the other half in the Social-Emotional condition first.

Task condition. The 20 groups under the Task condition worked on The Mined Road Problem, a situation test developed by the OSS Assessment Staff and modified for group discussion by Maier. The group was told to devise the best possible solution to a situation confronting five soldiers who must cross a mined road, using only miscellaneous debris found at the side of the road. The groups were allowed 50 minutes to arrive at their solution. A bonus of $25.00 was offered to the group with the best solution as judged by scoring criteria previously developed by Hoffman (1957).

Social-Emotional condition. The 20 groups under the Social-Emotional condition were asked to discuss how fraternities should handle the period of pledge training.[1] Their assignment was to arrive at a group statement of "the ideal policy and general program for handling pledge training in the houses." It was emphasized that there are no right or wrong answers to the question, but only various opinions regarding the ideal policy. Since each subject had recently undergone a long period of pledgeship prior to his initiation, and since each fraternity had its own unique training practices, a divergence of opinion among the subjects was expected. The groups were told that after the 50-minute discussion they would rate each fraternity represented in the group on how closely its program of pledgeship approximated the ideal program upon which the group had settled. Foreknowledge of this rating was introduced in an effort to increase the tendency on the part of each subject to defend the system his fraternity employed.

Final session questionnaire. After their second experience with group interaction the subjects were asked to contrast the two groups. The Final Session Questionnaire contained nine questions that served as a check on the effectiveness of the experimental manipulation, and specific predictions were made about the answer to each question. Subjects under the Task condition were expected to feel that (*a*) there were clearer grounds for deciding who was right or wrong and (*b*) there was greater need for intelligent, impersonal, and objective thinking. Under the Social-Emotional condition subjects were expected to feel (*a*) more concerned that their opinions be accepted, (*b*) more nervous or tense, (*c*) more annoyed or angry, (*d*) greater need for everyone to be sensitive to the feelings of others, (*e*) more judged or criticized as a person, (*f*) aware of greater competition for leadership, and (*g*) more bored.

Performance Measures

The subject's performance under the Task and Social-Emotional conditions was described by 35 measures of his overt behavior and informal status. Ten of the measures were derived from the observer's scoring of the act-by-act behavior of each subject; 16 of them were based on the peer ratings made by each subject of each other subject in his group after the 50-minute interaction was over; and 9 of them were observer ratings of each subject in the group, made after the interaction was over.

Observer scoring measures. The observer scoring system used Bales' (1950) interaction process categories in a modified form; some of Bales' 12 categories were collapsed to form a 7-category system. The 10 observer scoring measures were: the raw number of acts initiated in minutes 1–16, minutes 17–33, and minutes 34–50; and the percentage of total acts in Category 1, Agreement and Solidarity; Category 2, Laughing and Joking; Category 3, Suggestions; Category 4, Opinions and Orientations; Category 5, Questions; Category 6, Disagreement and Antagonism; and Category 7, Tension.

Peer rating measures. At the end of the 50-minute interaction each subject rated each other member of his group on eight informal status scales and eight behavior description scales, using a 10-point scale with no ties permitted. The description of the subject used in these analyses is the sum of the ratings received from the four other group members. The 16 peer rating measures were: Talking, Valuable ("... how valuable was he in guiding the discussion and getting things done?"), Liking, Leadership, Concern ("... how much concern had he for the feelings of others in the group?"), Influence ("... how much his suggestions were accepted by the other members of the group?"), Pleasant ("... how pleasant it was to be in the same group with him?"), Agreement (this and the seven behavior description ratings to follow differ only in the insertion of a different behavior in the question, "Regardless of how much he talked altogether, how would you rate each member of this group, excluding yourself, on how much he tended to agree with what others had said?"), Laughing and

[1] The period of "pledgeship" intervenes between the first and tentative acceptance of the potential fraternity member and his final acceptance and initiation.

Joking, Suggestions, Opinions and Orientations, Questions, Disagreement, Tension, Agressiveness, and Future Leader ("... how much would you like to see him as leader of this group, if the group were to meet again with a formal leader appointed?").

Observer rating measures. After the interaction was over, and while the group was completing its peer ratings, the experimenter rated each group member on nine scales, using a 10-point scale with no ties allowed. Although the ratings were made without reference to the scoring sheets, they were made by the same person who had scored the act-by-act behavior of the subjects in the group. The nine observer ratings were: Talking, Leader, Liking, Agreement, Laughing and Joking, Suggestions, Questions, Disagreement, and Tension.

RESULTS

Effectiveness of the Experimental Manipulation

Two comparisons between Task and Social-Emotional conditions may be made in an attempt to assess the effectiveness of the experimental manipulations. First, the Final Session Questionnaire may be examined to see whether the subjects perceived any differences between conditions, and, second, a comparison between conditions may be made in terms of the distribution of acts scored by the observer among the seven categories.

Table 1 presents the distribution of answers to the Final Session Questionnaire. Seven of the nine hypothesized differences between conditions were borne out by the data, and four of the seven differences in the predicted direction were statistically significant. The clearest difference between the two conditions is seen in the strong tendency to feel that the Task condition provided clearer grounds for deciding who was right or wrong as contrasted with the strong tendency to feel that there was greater need for sensitivity to other people's feelings in the Social-Emotional condition. Contrary to expectation, there was a clear tendency for subjects to report more anger in the Task condition.

Table 2 presents the distribution of overt behavior into the seven categories employed by the observer. Under the Task condition there was a stronger tendency to give suggestions and to disagree, whereas under the Social-Emotional condition there was a stronger tendency to laugh and tell jokes, to give opinions and orientations, and to show tension. With the exception of Category 5, Questions, the differences between conditions are significant at beyond the .001 level, using the z test for the difference between proportions.

The experimental manipulation of the group problem appears to have succeeded in creating reasonable approximations of task and social-

TABLE 1

RESPONSES TO THE FINAL SESSION QUESTIONNAIRE

Question	Number of subjects selecting each condition	
	Task	Social-Emotional
In which group did you feel more concerned that your opinions be accepted?	42	58
In which group did you feel more nervous or tense?	47	53
In which group did you feel more annoyed or angry?	59	41
In which group were there clearer grounds for deciding who was right or wrong?	75	25**
In which group was there the greater need for everyone to be sensitive to other people's feelings?	15	85**
In which group was there the greater need for intelligent, impersonal, and objective thinking?	49	51
In which group did you feel more that you were being judged or criticized as a person?	23	77**
In which group did people compete more for the leadership of the group?	49	51
In which group did you feel more bored with what was going on?	41	59*

* Deviation from 50-50 significant at .05 level, using one-tailed test.

** Deviation from 50-50 significant at .01 level, using one-tailed test.

TABLE 2

PERCENTAGE OF TOTAL ACTS IN EACH CATEGORY UNDER TASK AND SOCIAL-EMOTIONAL CONDITIONS

Category	Condition	
	Task	Social-Emotional
1. Agreement and Solidarity	13.0	15.0
2. Laughing and Joking	4.9	8.0
3. Suggestions	20.1	6.2
4. Opinions and Orientations	45.2	54.6
5. Questions	5.1	5.5
6. Disagreement and Antagonism	8.8	5.0
7. Tension	2.9	5.6
Total	100.0	99.9
Base number	14,901	16,550

emotional groups. Judging from the perceived clarity of evaluative standards, the greater frequency of attempted solutions, and the greater amount of disagreement, the group members under the Task condition were concerned primarily with arriving at a good solution, hopefully the best one. On the other hand, the subjects' perception that they are being judged, their concern that group members be sensitive to the feelings of others, the high level of overt tension, and, finally, the high level of laughing and joking within the group indicate that the social-emotional session was characterized by more persistent effort to solve the interpersonal problems generated by the inherent lack of consensus. It remains, then, to determine whether the underlying dimensions of performance remain the same under these two demonstrably different conditions.

Performance Factors

The 35 performance variables were intercorrelated separately for each condition, using product-moment correlations. Each 35 × 35 matrix was then subjected to a centroid multiple factor analysis.[4] Six factors were extracted from each matrix, the factors were rotated to an orthogonal solution by the normalized varimax criterion developed by Kaiser (1957).

Table 3 presents the first four factors of the rotated factor matrices under the two conditions. Under the Task condition the six factors account for 66.6% of the variance of the performance measures; under the Social-Emotional condition the six factors extracted account for 64.9% of the variance. For the purposes of this discussion only the first four factors in each matrix are presented since Factors VI and VII under both conditions account for little of the total variance of the measures and do not suggest any secure interpretation.

The first factor in both factor matrices, which may be labeled Task Prominence, occurs in much the same form across conditions. Table 4 shows the factor loadings of the variables with the 10 highest loadings under each condition. The variables with the highest loadings are the behavioral measures of total activity, giving opinions and orientations, and giving suggestions, together with the peer rating measures of total activity and task leadership (Leadership, Valuable, and Influence). One crude indication of the similarity between the Task Prominence factors under the two conditions is the fact that the rank order correlation between them is .90.

A second point of similarity between the factor structures under the two conditions is found in the examination of Factor III under the Task condition and Factor II under the Social-Emotional condition. All three peer rating measures of popularity (Liking, Concern, and Pleasant) are highly loaded on these factors, and they may be labeled the Likeability factor. However, the similarity between the Likeability factors under the two conditions lies only in the measures of popularity; several measures of the subject's behavior do not have similar loadings on these two factors. Table 5 shows the variables with loadings of greater than .40 on the Likeability factors. Under the Task condition, the three measures of laughing and joking all have loadings over .40 on the Likeability factor, whereas the highest loading for any of the three measures under the Social-Emotional condition is .16. The behavioral measures that do load the Likeability factor highly under the Social-Emotional condition are the various measures of agreement and disagreement. All but one of the six measures of the subject's tendency to agree or disagree with others have loadings greater than .40 on the Likeability factor under the Social-Emotional condition; none of the six measures has this high a loading under the Task condition. Finally, there is a slight tendency for the peer ratings on how much influence the subject had and how he was rated as a potential future leader of the group to be related to the Likeability factor more positively under the Social-Emotional condition.

The correspondence between Factor IV under the Task condition and Factor III under the Social-Emotional condition, although slight, deserves mention. As may be seen in Table 6 both factors have loadings greater than .40 on two of the three measures of tension, but only the high loading for the observer's scoring

[4] The correlation matrices and unrotated factor matrices may be found in Mann (1959). Copies of that work may be obtained from University Microfilm, Ann Arbor, Michigan.

TABLE 3
ROTATED FACTOR MATRICES UNDER TASK AND SOCIAL-EMOTIONAL CONDITIONS

Task Condition					Variable	Social-Emotional Condition				
Factor				h^2		Factor				h^2
I	II	III	IV			I	II	III	IV	
					Observer Scoring					
61	−03	18	−30	55	Acts first third	74	−04	−04	−11	58
75	01	08	−25	77	Acts second third	78	−05	−23	05	68
64	04	08	−29	60	Acts third third	87	−14	−09	10	81
−11	31	24	15	47	% Agreement	01	46	03	11	44
−40	−16	57	−12	54	% Laughing	−26	08	−06	57	41
19	14	−25	−50	37	% Suggestions	17	−15	03	−39	35
45	09	−10	07	26	% Opinions	12	10	−14	−14	64
−18	34	−11	67	62	% Questions	25	−06	−11	−32	45
07	−71	06	−01	52	% Disagreement	09	−64	26	−00	60
−29	−31	−12	59	58	% Tension	−46	−19	43	−18	51
					Peer Rating					
88	05	07	−37	93	Talking	93	05	−07	01	88
90	12	24	−21	92	Valuable	80	31	−13	−26	87
49	25	66	−08	77	Liking	42	69	05	23	71
92	05	15	−25	93	Leadership	90	21	−06	−12	89
15	24	71	18	64	Concern	07	70	−08	−12	54
78	08	20	−22	72	Influence	65	55	−10	−09	75
47	26	76	−06	87	Pleasant	38	72	−03	23	73
−46	23	39	43	73	Agreement	−12	69	−11	15	53
23	01	72	−00	65	Laughing	45	07	00	49	48
90	12	17	−09	86	Suggestions	73	32	−06	−22	76
93	14	20	−14	94	Opinions	79	02	−05	07	71
47	22	28	14	48	Questions	42	32	−07	−20	61
58	−19	−07	01	55	Disagreement	35	−45	19	04	78
−17	−06	−04	59	56	Tension	−50	−12	06	01	42
75	06	01	−15	78	Agressive	71	−26	12	14	81
67	33	35	−11	76	Future Leader	75	46	−10	−19	84
					Observer Rating					
73	09	01	−16	90	Talking	79	−07	−25	06	75
66	12	07	−16	82	Leader	78	08	−33	−05	81
34	64	24	−01	71	Liking	42	28	−64	09	68
05	72	26	01	60	% Agreement	24	39	−62	19	63
11	15	45	−13	46	% Laughing	35	16	−27	43	49
67	04	−01	−15	83	% Suggestions	73	−09	−20	−03	74
38	49	01	09	46	% Questions	49	06	−36	−05	64
52	−71	−25	30	73	% Disagreement	25	−55	38	−06	62
−10	−26	05	21	44	% Tension	−21	−33	63	−08	56

Note.—Decimal points have been omitted. The h^2 column is based upon the six unrotated factors.

of tension is maintained across conditions. The two factors may be labeled Tension in deference to their slight similarity, but the pattern of loadings is not consistent.

Finally, two factors for which no parallels appear may be noted. Factor II in the Task condition is highly loaded by three of the four observer measures of agreement and disagreement, as shown in Table 7, but the peer ratings on these behaviors are conspicuously missing. Factor IV in the Social-Emotional condition has three loadings over .40, as shown in Table 8, all of which are measures of laughing and joking. These two factors which are unmatched with a corresponding factor in the other condition represent the behaviors that fail to be related to the Likeability factor in that condition. Thus, in the task session, when Likeability is defined by the measures of popularity and by laughing and joking, some of the measures of agreement and disagreement define a separate factor. In the Social-Emotional con-

TABLE 4
THE TASK PROMINENCE FACTOR

Variable	Condition	
	Task	Social-Emotional
Leadership (P)	.92	.90
Talking (P)	.88	.93
Opinions (P)	.93	.79
Valuable (P)	.90	.80
Suggestions (P)	.90	.73
Acts second third (OS)	.75	.78
Talking (OR)	.73	.79
Acts third third (OS)	.64	.87
Suggestions (OR)	.67	.73
Aggressive (P)	.75	.71
Influence (P)	.78	.65
Leader (OR)	.66	.78
Acts first third (OS)	.61	.74

Note.—OS indicates an observer scoring measure, P a peer rating measure, and OR an observer rating measure.

TABLE 5
THE LIKEABILITY FACTOR

Variable	Condition	
	Task	Social-Emotional
Pleasant (P)	.76	.72
Concern (P)	.71	.70
Liking (P)	.66	.69
Laughing and Joking (P)	.72	.07
Laughing and Joking (OS)	.57	.08
Agreement (P)	.39	.69
Disagreement (OS)	.06	−.64
Disagreement (OR)	−.25	−.55
Influence (P)	.20	.55
Agreement (OS)	.24	.46
Future Leader (P)	.35	.46
Disagreement (P)	−.07	−.45

Note.—OS indicates an observer scoring measure, P a peer rating measure, and OR an observer rating measure.

dition, when the Likeability factor is defined by the same measures of popularity and the behavioral measures of agreement and disagreement, all three measures of laughing and joking define an independent factor.

DISCUSSION

One of the more striking findings of the present study is the high consistency with which a Task Prominence factor, closely resembling the one found by Slater (1955), emerges as the main factor of performance under both conditions. Despite considerable evidence that the two experimental conditions differ not only in the instructions given by the experimenter, but in the overt behavior and the subjective feelings of the group members, the Task Prominence factor remains remarkably constant across conditions. In each condition, behavioral measures of the total amount of participation combine with status measures of leadership and influence to define the factor.

TABLE 6
THE TENSION FACTOR

Variable	Condition	
	Task	Social-Emotional
Tension (OS)	.59	.43
Tension (OR)	.21	.63
Tension (P)	.59	.06
Liking (OR)	−.01	−.64
Agreement (OR)	.01	−.62
Questions (OS)	.67	−.11
Suggestions (OS)	−.50	.03
Agreement (P)	.43	−.11

Note.—OS indicates an observer scoring measure, P a peer rating measure, and OR an observer rating measure.

TABLE 7
THE AGREEMENT-DISAGREEMENT FACTOR

Variable	Condition
	Task
Agreement (OR)	.72
Disagreement (OS)	−.71
Disagreement (OR)	−.71
Liking (OR)	.64

Note.—OS indicates an observer scoring measure, OR an observer rating measure.

TABLE 8
THE LAUGHING AND JOKING FACTOR

Variable	Condition
	Social-Emotional
Laughing and Joking (OS)	.57
Laughing and Joking (P)	.49
Laughing and Joking (OR)	.43

Note.—OS indicates an observer scoring measure, P a peer rating measure, and OR an observer rating measure.

The Task Prominence factors isolated in the present study and in Slater's (1955) research combine the Activity and Task Ability factors described by Carter (1954) and Bales (1956) as independent dimensions. This discrepancy deserves further attention. The experimental conditions that differentiate this and Slater's study from studies in which the Activity and Task Ability factors do not fuse offer some clues to understanding the conditions under which they tend to fuse. Slater's study and the present one are similar in inves-

tigating individual performance in *ad hoc* groups whose members were, by and large, unacquainted with one another prior to the experimental situation. In contrast, most of the studies that report independent Activity and Task Ability factors employed natural groups in which there was considerably more history to the group before the measures were taken. Perhaps a gradual differentiation takes place over time between the two aspects of Task Prominence, namely, Activity and Task Ability. This possibility derives some support from Slater's finding that a gradual differentiation does take place between the *most* active and the *most* valuable member of the group, since most active and most valuable correspond closely to the Activity and Task Ability factors isolated by Carter.

While the sociometric measures of popularity define an independent factor under both Task and Social-Emotional conditions in the present study, the behavioral measures that share in the definition of the factors vary markedly from one condition to the other. Laughing and joking behavior is highly related to Likeability only under the Task condition, whereas agreement and the avoidance of disagreement are strongly related to Likeability only under the Social-Emotional condition. These findings both replicate and challenge the results of previous research.

In light of the differences between the conditions that may be inferred from the responses to the Final Session Questionnaire, these findings suggest a number of conclusions about the dynamics of popularity. When the group is engaged in a formal, intellectual task, one in which there is both competition with other groups and an implied standard of excellence, the persons who will be liked appear to be those who can break through the tension by laughing and joking. Such behavior may also serve to reduce the external threat, represented by a difficult intellectual task, allowing the group members to maintain the spirit of a lighthearted game, rather than a test of their intelligence. In the Task condition most of the jokes tended to be in the form of ridiculous proposals for the solution to The Mined Road Problem, proposals offered in jest and received with laughter. Turning to the independence of variables concerning agreement and disagreement from the Likeability factor, under the Task condition, perhaps the most sensible approach to this finding would be to view the impact of agreement or disagreement in the light of the task contributions to which they are directed. When the content of the discussion is focused on an external task, like The Mined Road Problem, reactions of approval or disapproval apparently do not tend to alter significantly the affective or liking relationships in the group. Although these data offer no direct test of the view, the measures of agreement and disagreement may be independent of Likeability because such a large proportion of the task acts to which they are addressed are merely statements of fact about The Mined Road Problem, or the entire problem may have remained sufficiently distant from the fundamental ego involvements of the subjects to evoke only superficial involvement in their proposals for its solution.

The clearest differences between the Task and Social-Emotional conditions, the higher level of tension and the greater sense of personal threat and criticism under the Social-Emotional condition, offer some leverage for understanding how individual behavior relates to popularity in the Social-Emotional condition. In the discussion of fraternity life, the group members, all of whom were neophytes of different fraternities, seem to have been highly involved in the content of the discussion. Under such conditions, the person who showed agreement and avoided disagreement with others was likely to be well liked. However, as the discussion increased in personal relevance, the danger that laughing and joking would provoke resentment and tension increased correspondingly. Whereas in the Task condition the jokes were directed out of the group toward the experimenter and his frustrating intellectual problem, in the Social-Emotional condition the jokes were directed within the group toward the activities of the fraternities represented around the table. Very likely a sufficient number of these jokes, perhaps even ones that evoked laughter from the group, carried with them a note of scorn either for some other fraternity and its ways or for the whole fraternity system.

Summing up these speculations about the Likeability factor, these data suggest that the way in which overt behavior relates to popularity varies with the extent to which the group is task- or social-emotionally oriented. As the tension and integrative needs of the group increase, agreement and the avoidance of disagreement become important to the group members, and liking tends to be given to those who show the most support for others' positions; at the same time, laughing and joking do not produce liking in others, as they do in the task oriented groups, because the target of the jokes shifts from outside the group to inside, where feelings are considerably more sensitive.

In conclusion, the systematic study of performance factors as they vary under Task and Social-Emotional conditions challenges the three-factor proposals of Carter (1954) and Bales (1956) in several ways. The factors of Activity and Task Ability that they propose fuse into one Task Prominence factor in both conditions, a result that suggests the need for more exact knowledge about how Activity and Task Ability come to be differentiated from each other. Their Likeability factor turns out to be stable over both experimental conditions only with regard to the postmeeting measures of informal status; the behavioral measures that share in the definition of this factor vary markedly across conditions. Further consideration of these behavioral variables may extend our conception of the Likeability factor beyond its present bounds. Finally, the three-factor proposals of Carter and Bales fail to include a Tension factor, which appeared in both conditions, and the behavioral factors of Agreement-Disagreement and Laughing and Joking, which emerge as independent factors when the behaviors involved are not related to the Likeability factor. It might be wiser to avoid premature closure on any set of performance factors until more is known about the effect that variation in the situation, the size and nature of the group, or other experimental procedures has upon the resultant factor structure. The present study suggests some of the outlines of the more complex view of performance factors that may emerge from such needed research.

SUMMARY

The proposal by Carter (1954) and Bales (1956) that three major factors account for much of the variance in measures of performance in small groups was subjected to further test by obtaining factor structures from performance measures under two demonstrably different experimental conditions. In the Task condition the group was oriented to an external problem, an intellectual task, and in the Social-Emotional condition a topic of high personal relevance created problems of integration and tension reduction. Twenty five-man groups were run under each condition. The performance measures taken on the 100 subjects were factor analyzed separately for each condition.

The results of the two factor analyses revealed a Task Prominence factor that was highly stable across conditions; it was identified as a fusion of the Activity and Task Ability factors proposed by Carter (1954) and Bales (1956). A factor, labeled Likeability, showed consistent loadings from postmeeting measures of popularity, but highly inconsistent loadings from the behavioral measures that shared in its definition. Some implications of the results for a more complex interpretation of Likeability and the dynamics of popularity than the one suggested by Carter and Bales were presented. A factor of low stability, labeled Tension, was extracted under each condition. Finally, independent behavior factors were formed by measures of either agreement and disagreement or laughing and joking, depending on which behavior was not related to Likeability under that condition.

REFERENCES

BALES, R. F. *Interaction process analysis: A method for the study of small groups.* Cambridge, Mass.: Addison-Wesley, 1950.

BALES, R. F. Task status and likeability as a function of talking and listening in decision-making groups. In L. D. White (Ed.), *The state of the social sciences.* Chicago: Univer. Chicago Press, 1956. Pp. 148–161.

BORGATTA, E. F., COTTRELL, L. S., & MANN, J. H. The spectrum of individual interaction characteristics: An inter-dimensional analysis. *Psychol. Rep.*, 1958, **4**, 279–319.

CARTER, L. F. Recording and evaluating the performance of individuals as members of small groups. *Personnel Psychol.*, 1954, **7**, 477–484.

HOFFMAN, L. R. Homogeneity of member personality

and its effect on group problem-solving. Unpublished doctoral dissertation, University of Michigan, 1957.

KAISER, H. V. The varimax method of factor analysis. Unpublished doctoral dissertation, University of California, 1957.

MAIER, N. R. F. *Principles of human relations*. New York: Wiley, 1952.

MANN, R. D. The relation between personality characteristics and individual performance in small groups. Unpublished doctoral dissertation, University of Michigan, 1959.

PARSONS, T. *The social system*. Glencoe, Ill.: Free Press, 1951.

SLATER, P. E. Psychological factors in role specialization. Unpublished doctoral dissertation, Harvard University, 1955.

(Received June 20, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 683-685

CRITIQUE AND NOTES

SEX, ETHNICITY, AND PLAY PREFERENCES OF PRESCHOOL CHILDREN

BOYD R. McCANDLESS AND JUNE M. HOYT[1]
Iowa Child Welfare Research Station

In the literature concerned with ethnocentrism, few studies have been concerned with preschool age children. Moreno (1934) and Criswell (1937), using sociometric techniques, found no evidence of ethnocentric cleavage until about 8–10 years of age. Lambert and Taguchi (1956) suggest that such cleavage exists in preschool children, but is not demonstrated in the kinds of choice required by the standard sociometric techniques. They devised choices that they assumed were more meaningful to young children, and found that their measure elicited ethnocentric choices in Oriental children in a Montreal preschool, but that the comparable trend among Occidental children was not statistically significant. Among possible explanations, the authors considered that the Oriental children, as a minority group, perceive more threat in a mixed racial situation than do the Occidentals, and hence cling more closely together.

In an intensive study of race awareness in Negro and white children of preschool age, Goodman (1952) found that cross-racial interactions of children did not differ in frequency from that which would be expected by chance, although interactions of white with white were 8% below chance, while those of Negro with Negro were 9% above chance. Goodman concludes:

> The high rate of Negro-Negro interaction and the low rate of white-white interaction must therefore result from the operation of variables of a nonracial order. We conclude that these variables are personality traits which are unequally distributed between the Negro and the white children in this particular group (p. 259).

In the present study it was hypothesized that, in a situation where social mixing of races is common, preschool children show little tendency toward ethnic cleavage. Choice of play companions in a free play situation was assumed to be an interaction, meaningful to the child, in which this hypothesis could be tested.

[1] The authors are particularly indebted to Hannah Lou Bennett and Carolyn B. Bilous for providing the opportunity to conduct this study; to Betty Crooker and Mary Rushit for opening their preschool groups to the senior author; and to Mark Beals for his help in establishing reliability of observations. The basic data were collected while the senior author was a Carnegie Visiting Professor at the University of Hawaii.

METHOD

Subjects were 33 children of Oriental (non-*haole*) ethnic origin and 26 Caucasian (*haole*) children at the University of Hawaii Preschools. The age range was from 3 years 6 months to 5 years 3 months, with a mean age of 4 years 5 months. All except one child had been in the preschools for at least 6 months, this exception being the senior author's youngest child, who was a newcomer to her group. There were 17 boys and 16 girls in the non-*haole* group and 11 boys and 15 girls in the *haole* group. Three subjects were eliminated: one girl because she did not fall into either of the ethnic groups under study, one blind child who did not socialize with anyone, and one mentally defective boy.

The children were observed in free play in their regular preschool groups, the ethnic and sex composition of the three groups being approximately the same. The time sampling method employed in making observations consisted of 15 5-minute observations for each child, during which time the duration and type of his interactions with other children were recorded. Three observers studied the children over a period of approximately 3.5 months. Percentage of exact agreement for Observers A and B was 92; and for A and C, 93. A tabulation was made of total time spent with every child with whom the child under observation interacted during the observation periods, and total time was computed for interactions with children in each of the two ethnic groups.

The chance expectancy for time spent playing with *haole* and non-*haole* children was computed for each preschool group on the basis of the ethnic proportion in the group. For example, if a group is made up of 11 *haole* children and 10 non-*haole* children, the chance expectancy for a given *haole* child for each ethnic group is 50% of his time. If he spent a total of 60 minutes in free play, it would be expected that 30 of these were spent with other *haole* children, and 30 with non-*haole* children. From this base, discrepancy scores were computed. For our hypothetical *haole* child, if he spent 45 minutes playing with other *haoles*, the discrepancy score would be +15. The same procedure was used to study free play as related to sex.

Comparisons were made, using a chi square analysis, of non-*haole* and *haole* children's choices of play companions based on this discrepancy score. The mean discrepancy was determined for each ethnic group, and separate comparisons were made within each group by sex of child. Comparable analyses were made for choice of companion by sex without regard to ethnic group.

RESULTS

The chi square analysis, Yates' correction,

TABLE 1

MEAN TIME AND STANDARD DEVIATIONS IN MINUTES SPENT AND PREDICTED BY CHANCE BY DESIGNATED GROUPS IN PLAY WITH EACH OTHER

	Actual Time		Chance Expectancy Time	
	M	*SD*	*M*	*SD*
Boys with boys	114.6	42.9	72.9	22.8
Boys with girls	46.4	30.1	88.1	23.1
Girls with boys	42.8	27.2	63.3	28.8
Girls with girls	82.7	33.8	62.1	25.0
Orientals (non-*haole*) with non-*haole*	91.1	49.6	78.6	34.6
Non-*haole* with Caucasian (*haole*)	53.8	24.7	66.2	26.3
Haole with *haole*	63.4	26.6	56.8	15.6
Haole with non-*haole*	75.7	39.2	82.3	27.7

TABLE 2

MEAN TIME IN MINUTES BEYOND CHANCE EXPECTANCY SPENT IN OBSERVED PLAY WITH COMPANIONS INDICATED

Play Companions	Mean Excess over Chance	*SD*	*t*
Without regard to ethnic group:			
Boys playing with boys	41.7	29.3	
Girls playing with girls	20.5	15.9	
Total difference from expectancy (boys plus girls)	30.6	25.5	10.10*
Difference between boys and girls	21.2		3.44*
Without regard to sex:			
Non-*haoles* playing with non-*haoles*	12.4	24.5	
Haoles playing with *haoles*	6.6	25.3	
Total difference from expectancy (boys plus girls)	9.2	25.11	2.87*
Difference between non-*haoles* and *haoles*	5.8		.88
Ethnic group by sex:			
Non-*haoles* playing with non-*haoles*			
Boys	15.4	28.3	
Girls	9.3	18.3	
Difference	6.1		.70
Haoles playing with *haoles*			
Boys	11.9	33.2	
Girls	2.7	16.3	
Difference	9.2		.88

* $p = <.01$ for 57 *df*.

demonstrates that, disregarding ethnic group, boys overchoose boys as play companions, and girls overchoose girls ($\chi^2 = 32.9$; for $p = .01$, 1 *df*, 6.635 required). There was also an ethnic cleavage in choice of play companion (χ^2 for 1 *df* = 6.9). In other words, boys played more with boys, girls with girls, disregarding ethnic group; and, disregarding sex, *haoles* played more with *haoles*, non-*haoles* with non-*haoles*.

Table 1 summarizes the distribution of play times by subgroup., and Table 2 shows the mean time beyond chance expectancies associated with these choices. The sex cleavage was significantly more extreme for boys than girls, although both spent more than the expected proportion of their time playing with children of their own sex. The ethnic cleavages did not differ significantly for the *haole* and non-*haole* children, contrary to the finding of Lambert and Taguchi (1956), nor did boys differ from girls in this respect.

DISCUSSION

Contrary to hypothesis, it was found that children in this study differentially selected play companions of their own ethnic group, even in an atmosphere that, subjectively, would be presumed to be more accepting of interracial companionships than might be expected in the Montreal setting of the Lambert and Taguchi (1956) study.

The statistical analysis revealed a trend that had not been suggested to the senior author as he gathered the rather intensive observational material for the children. At no time did he hear an instance of verbal rejection that included ethnic reference, and his subjective judgment was that friendships and play contacts were the usual functions, found in ethnically more homogeneous preschool groups, of personality attributes, play interests, sex, and accident. In other words, he would have predicted the sex difference, but not the ethnic difference.

The authors consider that prejudice is based on hostility, and, for this reason, tentatively suggest that the ethnic "discrimination" found among these children is not prejudice as usually thought of, but some sort of "comfort" differential by ethnic group. Ethnically different background, in the sense of such factors as greater commonality of food habits within the ethnic group, colloquialisms drawn from the original parent language (such as the terms used for toileting), perhaps subtly different play experiences and interests, may operate to make the *haole* youngster feel more comfortable with other *haoles* (he has "more in common with them"). Similar factors may operate for the non-*haoles*. In other words, children in one ethnic group do not actively *avoid* members of the other group, but differentially *approach* the members of their own group.

Informal checking over the families and their patterns of out-of-school social interaction indicated that neither the non-*haole* nor *haole* parents in this Honolulu metropolitan area associated selectively with each other (i.e., the *haole* parents knew the non-*haole* parents as well as they did the *haole*, and did not appear to have selective out-of-school association with them by ethnic group; and

vice versa). It is probable, however, even in cosmopolitan Honolulu, that, over-all, the out-of-school play and visiting contacts of non-*haole* youngsters are more frequently with other non-*haoles*; and likewise for the *haole* children. This probably contributes to the "comfort" or "compatibility" factor previously mentioned.

SUMMARY

Preschool children of Oriental (non-*haole*) and Caucasian (*haole*) background were observed, using a time-sampling technique; and their choices of play companions were analyzed. Children in both ethnic groups were found to prefer play companions of their own sex, this tendency being more pronounced in boys than in girls. *Haole* children spent more of their time with *haole* children and non-*haole* children with non-*haole* children than would be predicted on the basis of proportional representation of each of the two groups. The subjective impressions of the senior author were that this cleavage was not due to prejudice in any conventional sense, but possibly to differential "comfort" due to more common backgrounds among the members of a given ethnic group.

REFERENCES

CRISWELL, JOAN H. Racial cleavage in Negro-white groups. *Sociometry*, 1937, **1**, 85–89.

GOODMAN, MARY E. *Race awareness in young children*. Cambridge, Mass.: Addison-Wesley, 1952.

LAMBERT, W. E., & TAGUCHI, YOSH. Ethnic cleavage among young children. *J. abnorm. soc. Psychol.*, 1956, **53**, 380–382.

MORENO, J. L. *Who shall survive?* Washington: Nervous and Mental Disease Publ. Co., 1934.

(Received April 8, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 686-687

A COMPARISON OF MENTAL RETARDATES AND NORMALS ON KINESTHETIC FIGURAL AFTEREFFECTS

HERMAN H. SPITZ AND RONALD S. LIPMAN

Edward R. Johnstone Training and Research Center, Bordentown, New Jersey

A previous experiment (Spitz & Blackman, 1959) indicated that mentally retarded males exhibit a poorer capacity for visual satiation, as measured by a visual aftereffect (VAE) test, than do normal males of approximately equal CA. Since there is no correlation of individual scores on VAE and kinesthetic aftereffect (KAE) tests, at least with college students (Spitz & Lipman, 1960), the question is posed of whether the group differences found in visual satiation also hold in kinesthetic satiation (Köhler & Dinnerstein, 1947). Furthermore, use of a KAE test provides a control for the possibility that the results of the previous study were due in part to the retardates' inability to fixate steadily. Female subjects, not included in the previous experiment, were included in the present study.

METHOD

As shown in Table 1, the average CA of the retarded group, all residents at a state training school, approximated the average CA of a comparison group of college sophomores.[1]

The KAE test apparatus was similar to the one described by Klein and Krech (1952). On the subject's right was the variable test (T) block on which the position of a rider could be read to the nearest $\frac{1}{16}$ inch. The comparison T block on the subject's left was 1.5 inches wide and could be quickly replaced by the Inspection (I) block, which was 2.5 inches wide.

The task was explained to the subject, who stood between the T blocks, gripping them with the forefinger and thumb of his left and right hands. Starting from random positions alternately below and above the point of objective equality (POE), the subject found the place on the right hand variable T block which felt as wide as the left hand comparison T block. He was given four "feedback" practice trials. That is, after each judgment, his hand was placed at POE. This procedure was introduced in an attempt to reduce variability, particularly in the retarded group. Next, he was asked to make four more equality judgments, this time without feedback. These judgments were averaged to yield a Control score. Then, with the same two fingers of his left hand, he rubbed the larger I block for 1.5 minutes, in approximate time to a metronome set at 108 beats per minute, after which the original T block was quickly returned for the subject to make four post-inspection equality judgments. The average of these four post-inspection judgments subtracted from the Control score yielded a Satiation score for each subject.

During the experiment, subjects wore goggles which blocked the lower half of the visual field. Whereas the normal subjects put on the goggles before the practice feedback trials, the retardates were not asked to wear them until after the feedback trials. Care was taken, however, to keep them from looking down at their judgments. In any case, all subjects wore the goggles for the Control Measurements and until the experiment ended.

The reliability of individual scores on the KAE test used here has been shown to be significant (Lipman & Spitz, 1959; Spitz & Lipman, 1960), although the reliability coefficient is much lower for retardates than for normals (.34 as compared to .74).

RESULTS AND DISCUSSION

Since there were no significant sex differences, within groups, in mean Satiation scores, the combined group of normals was compared with the combined group of retardates. As shown in Table 2, the mean Satiation score of the normals was significantly larger than that of the retardates, confirming the previous finding with VAE. However, whereas on the VAE test, retardates were significantly more variable than normals, no such differences in variability were found in the present study.

Although, as noted, sex differences within the groups were not significant, the females of both groups had lower Satiation scores than the respective males in line with the trend reported by Wertheimer (1954a, 1954b) and by Spitz and Lipman (1960). On the other hand, the present results may be considered inconsistent with a previous finding that female retardates perceive spiral aftereffects more readily than retarded males or, in fact, than normal females (Spitz & Lipman, 1959). The spiral, however, is more open to subjective distortion in reporting results and, in any case, may not be the most adequate measure of satiation.

It should be noted, also, that on tests of figural aftereffects retardates perform more like the brain damaged subjects reported by Spivack and Levine (1959) than those of the Klein and Krech study (1952).

In summary, the present results with KAE, when added to previous results with VAE, indicate

[1] The authors are most grateful to Marshall P. Smith, Chairman of the Psychology Department of Trenton State College, for his cooperation during this study.

TABLE 1

CHRONOLOGICAL AGE DISTRIBUTION OF RETARDED AND NORMAL GROUPS, WITH IQ DISTRIBUTION OF RETARDED GROUPS

	N	Mean Age (yr.)	Range	Mean IQ	Range
Retarded Males	35	19.30	17.08–22.42	70.34	56–86
Normal Males	32	20.28	18.83–22.67		
Retarded Females	20	18.83	15.92–21.83	70.90	57–80
Normal Females	50	19.58	18.58–20.75		

TABLE 2

DIFFERENCES BETWEEN MEAN SATIATION SCORES OF NORMALS AND RETARDATES

	N	*M*	*SD*	*t*	*p*
Normals	82	1.59	1.95	2.18	$<.05$
Retardates	55	.85	1.91		

that mentally retarded subjects are characterized by weaker satiation processes than are normals of comparable CA.

REFERENCES

KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain injured. *J. Pers.*, 1952, **21**, 118–148.

KÖHLER, W., & DINNERSTEIN, DOROTHY. Figural after-effects in kinesthesis. In *Miscellanea psychologica Albert Michotte*. Paris: Librairie Philosophique, 1947. Pp. 196–220.

LIPMAN, R. S. & SPITZ, H. H. An experimental test of the similarity between Hull's inhibition construct and Köhler's satiation construct. Paper read at Eastern Psychological Association, Atlantic City, April 1959.

SPITZ, H. H., & BLACKMAN, L. S. A comparison of mental retardates and normals on visual figural aftereffects and reversible figures. *J. abnorm. soc. Psychol.*, 1959, **58**, 105–110.

SPITZ, H. H., & LIPMAN, R. S. Some parameters in the perception of the spiral aftereffect. *Percept. mot. Skills*, 1959, **9**, 81.

SPITZ, H. H., & LIPMAN, R. S. The reliability and intercorrelation of individual differences on visual and kinesthetic figural aftereffects. *Percept. mot. Skills.*, 1960, **10**, 159–166.

SPIVACK, G., & LEVINE, M. Spiral aftereffect and measures of satiation in brain-injured and normal subjects. *J. Pers.*, 1959, **27**, 211–227.

WERTHEIMER, M. Constant errors in the measurement of figural aftereffects. *Amer. J. Psychol.*, 1954, **67**, 543–546. (a)

WERTHEIMER, M. The differential satiability of schizophrenic and normal subjects: A test of a deduction from the theory of figural after-effects. *J. gen. Psychol.*, 1954, **51**, 291–299. (b)

(Received March 3, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 688-689

OPERANT CONDITIONING OF VERBAL BEHAVIOR OF DYSTHYMICS AND HYSTERICS

DONNA M. GELFAND AND C. L. WINDER[1]
Stanford University

EYSENCK (1955) has postulated that neurotics of the dysthymic type should condition more readily than neurotics of the hysteric type. This hypothesis was given support by Franks (1956) in a study of GSR and eyeblink conditioning. Franks concluded, in addition, that conditionability is related to extraversion-introversion. The present experiment investigated the possible relationship between conditionability of verbal behavior and type of psychopathology. The hypotheses are that dysthymics condition more readily and stably and therefore make more conditioned responses in an acquisition and a counterconditioning series than do hysterics.

METHOD

The subjects consisted of 26 female inpatients at a state mental hospital. Ten subjects fell into the dysthymic category, while 16 were classified as hysterics. Hysteric subjects were described clinically as having one or more of the following characteristics: conversion symptoms, hysteria, psychopathic personality, or some similar type of character disorder. Hysterics tended to exhibit histories of poor impulse control and antisocial behavior. Dysthymic subjects displayed one or more of the following primary characteristics: manifest anxiety, reactive depression, or obsessive-compulsive features. Two judges, independently categorizing the subjects as dysthymic or hysteric, agreed on 22 of 25 cases (exact chi square, $p = .003$).

Guilford's R scale (1940) was used in this study as a measure of extraversion. The Taylor Manifest Anxiety (*MA*) scale, MMPI scales *L*, *F*, *K*, and Edwards (1957) Social Desirability scale were also administered. The personality inventories were given prior to the conditioning procedure.

One hundred sixty commonly used verbs were randomly selected, and each was typed on a separate 3 × 5 white index card. All verbs were in the past tense. Six pronouns—I, We, He, They, She, and You—were placed below each verb with their respective orders randomized. The subject was shown a sample card and instructed to make up a sentence containing the verb indicated on the card and beginning with any one of the pronouns. The subjects was asked to construct one such sentence for each of the 160 cards.

Operant level of responding was determined in the first 20 trials. In the following 60 acquisition trials, the experimenter said, "Good," in a flat, unemotional tone at the end of any sentence which the subject began with I or We. The last series of 80 card presentations consituted the extinction or counterconditioning phase, during which reinforcement was shifted to He and They responses. No subject reported awareness of the purpose of the experiment.

RESULTS

The two groups were not significantly different in operant level of responding as measured by the number of first person pronouns given in the first 20 nonreinforced trials. In the 60-trial acquisition series, dysthymics produced significantly more first person pronouns than hysterics ($t = 1.98$, $p = .05$, one-tailed test). During the extinction series, the dysthymic group continued to give more first person responses than hysterics ($t = 1.71$, $p = .05$, one-tailed test). Hysterics obtained significantly higher scores on the Edwards Social Desirability scale than dysthymics ($t = 2.20$, $p = .05$, two-tailed test). The samples did not differ significantly on the Guilford R scale, the *MA*, or MMPI scales *L*, *F*, and *K*.

DISCUSSION

Results of this study indicate it is possible, on the basis of symptomatology, to choose groups which differ in conditionability of verbal behavior. However, it is not yet possible to define, in terms of scores on psychometric tests, the relevant psychological variables upon which such groups differ. The findings do not support the relationship between extraversion and resistance to conditioning postulated by Franks (1956) and Eysenck (1955). The only variable, other than conditionability, on which the groups differ is social desirability. There is, however, no apparent relationship between ease of conditioning and social desirability ($r = -.054$).

Other investigators have found relationships between conditionability of verbal behavior and such variables as compliance (Sarason, 1957) and dependency-anxiety (Cairns, 1959). Investigation of possible relationships between these variables and symptomatology characteristic of dysthymics and hysterics seems important for future research.

[1] This study, done at Agnews State Hospital, Agnews, California, was possible only because of the warm cooperation of the hospital staff. Jerry S. Wiggins was very helpful with suggestions regarding the investigation.

REFERENCES

CAIRNS, R. B. The influence of dependency-anxiety on the effectiveness of social reinforcers. Unpublished doctoral dissertation, Stanford University, 1959.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–151.

GUILFORD, J. P. *Inventory of factors STDCR.* Beverly Hills, Calif.: Sheridan Supply, 1940.

SARASON, I. G. Interrelations among individuals difference variables, behavior in psychotherapy, and verbal conditioning. Paper read at Western Psychological Association, Eugene, Oregon, May 1957.

(Received March 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 690-692

PERCEPTUAL SELECTIVITY, MEMORY, AND ANTI-SEMITISM

LEE PULOS AND BERNARD SPILKA

Larue Carter Memorial Hospital, Indianapolis, Indiana *University of Denver*

THAT persons who score as anti-Semitic on objective scales are able to discriminate fairly accurately between photographs of Jews and non-Jews has been demonstrated by Allport and Kramer (1946), and Lindzey and Rogolsky (1950). They also noted that persons scoring as anti-Semitic tended to identify more photographs as Jewish than those who scored low in anti-Semitism. Their findings, however, have been extensively qualified and questioned on the basis of group composition of the photographs and experimental controls (Carter, 1948; Elliot & Wittenberg, 1955; Scodel & Austrin, 1957).

Seeleman (1940) noted that those who indicated unfavorable attitudes toward Negroes were significantly less able to discriminate among a group of photographs of Negroes than those with positive attitudes.

Consideration of these studies suggests a need for further evaluation of the hypothesis relating perceptual accuracy to prejudice, plus examination of the additional hypothesis that once the in-group–out-group distinction is made, there is no further need to distinguish among members of the out-group.

The present investigation was thus conducted to verify first, the relationship between anti-Semitism and perceptual accuracy, and second, the possibility of differential memory for persons scoring high and low in anti-Semitism on photographs initially identified as Jewish or non-Jewish. In the latter case, Seeleman's (1940) study suggests that a group scoring high in anti-Semitism, if they are able to identify accurately photographs of Jews and non-Jews, should be less able than persons scoring low in anti-Semitism to remember those photographs defined initially as Jewish.

METHOD

Subjects

Two hundred and twenty male undergraduate students at the University of Denver served as subjects. Following administration of the anti-Semitism scale of Levinson and Sanford (1944), the 40 students who scored highest and the 40 who scored lowest on this scale were selected for further study. The range of scores for the entire 220 subjects was from 52–286. The mean of the scores for the low anti-Semitic group was 68 and for the high group, 208. Both groups were equivalent in age, the low group having a mean of 21.2 years and the high group a mean of 21.1 years.

Materials

The photographs to be identified were selected from yearbooks of the University of Denver. The religious affiliation and ethnic identity of each person was established by checking personal history and religious preference records in the Registrar's office. All pictures were of male seniors in cap and gown. This procedure was a departure from earlier studies and was employed to restrict possible cues employed in identification of Jews and non-Jews, and also to provide an element of uniformity for all the photographs presented. Seven judges rated each photograph on a five-point scale that represented the degrees of Jewish and non-Jewish appearance of the photographs. Scale values were computed for each photograph, and all photographs that were consistently rated as extreme or were otherwise easily distinguished and remembered were removed. Of the original 82 photographs employed, 20 photographs of Jewish and 20 of non-Jewish persons were selected. These were then broken into three sets as follows:

Set I: Mixed Jewish and non-Jewish (10 of each matched as to scale value).

Set II: New set of Jewish faces (10) plus the Jewish faces employed in Set I. (Photographs in both sets were matched individually in scale values.)

Set III: New set of non-Jewish faces (10) plus the non-Jewish faces employed in Set I. These were also individually matched in scale value as in Set II.

Procedure

Following the selection of the 40 high and 40 low scoring subjects, each subject was individually contacted and tested on the three sets of photographs. The order of presentation of all pictures in Set I was carefully arranged so that sharp contrasts between faces and scale values could not be used as cues to distinguish the Jewish from the non-Jewish photographs. The order of presentation for Sets II and III was random as defined by reference to a table of random numbers. Following administration of Set I, a counterbalancing design was used and half of the subjects in each group were given Set II first, then Set III. The remaining subjects received Set III first, then Set II.

The initial task was to define each photograph in Set I as Jewish or non-Jewish. Each photograph was exposed for 5 seconds. Following administration of Set I, a 4-minute block design task was given to each subject. Sets II and III were then administered to each subject, who was then requested to identify the photographs that had been included in Set I.

RESULTS AND DISCUSSION

Table 1 presents the comparative data for the high and low groups on both identification and

TABLE 1

SUMMARY OF ACCURACY IN IDENTIFICATION AND MEMORY FOR LOW AND HIGH ANTI-SEMITIC GROUPS

Group	Low anti-Semitic (N = 40)		High anti-Semitic (N = 40)	
	M	SD	M	SD
Correct identifications of Jews	4.1	1.76	6.1	1.49
Non-Jewish pictures identified as Jewish	1.4	1.09	2.1	1.46
Pictures identified as Jewish	5.5	1.47	8.2	1.48
Jewish pictures that were accurately identified in Set I and were again accurately identified in Set II	3.2	1.46	4.9	1.53
Non-Jewish pictures that were accurately identified in Set I and were again accurately identified in Set III	6.2	1.45	5.4	1.73

memory tasks. An analysis of the total number of pictures identified as Jewish revealed that the High A-S group was significantly more accurate than the Low A-S group in identification of Jewish photographs ($p < .01$). The difference between the two groups in accuracy of identification of the non-Jewish photographs fell short of significance. The High A-S group identified significantly more photographs as Jewish than did the Low A-S group ($p < .02$), but an analysis of covariance revealed that the accuracy in identification of Jewish photographs holds when the total number of Jewish identifications of non-Jews is controlled for ($p < .01$). The correlation between the accuracy scores and frequency of identifying the photographs as Jewish was .40 ($p < .01$, 78 df) thus necessitating the use of the above procedure.

In order to compare the high and low groups in regard to accuracy of memory for the Jewish and non-Jewish faces, an analysis of variance was computed for those pictures which were correctly identified in Set I and then again correctly identified in Sets II and III.

Table 2 indicates a significant interaction between the groups and pictures, which necessitated a further comparison of the subgroups means through the use of the Tukey process (Walker & Lev, 1953). When the accuracy means (Table 1) were corrected for the group and photograph effects, the High A-S group obtained a mean accuracy score of 5.55 on the Jewish photographs and 4.30 on the non-Jewish photographs. The Low A-S group received an accuracy score of 4.29 Jewish identifications and 5.54 for the non-Jewish identifications.

In order for these corrected means to be significant at the .01 level, a difference of .38 between any two means would have to be demonstrated. It may be seen that a difference of 1.25 obtained between the photographs within each of the groups. Apparently the High A-S group remembered significantly more of the Jewish faces than did the Low A-S group, however the latter group demonstrated more accuracy in memory for the non-Jewish faces.

The results of this investigation relating to identification of in-out-group members are clearly supportive of the findings of Allport and Kramer (1946) and of Lindzey and Rogolsky (1950). The second hypothesis relating to memory is contradicted by the data and suggests the operation of different processes in memory for Jews than were suggested by Seeleman (1940) in her study of memory for Negroes. The findings would initially suggest that the High A-S group remembered the Jewish photographs significantly more accurately than the non-Jewish photographs, in addition to demonstrating a significantly higher degree of accuracy than the Low A-S group in memory of the Jewish photographs. The situation, however, is reversed for the non-Jewish photographs and this was not expected. Both the High and Low A-S groups appear to show differential sensitivity to in-out-group distinctions.

SUMMARY AND CONCLUSIONS

The present study was conducted to verify first the apparent tendency of anti-Semitic persons to identify Jews more accurately than do non-anti-Semites. Second, it was hypothesized that once an anti-Semitic person has distinguished between members of the in-group and the Jewish out-group, he should have no further need to distinguish among the out-group members, and his memory for those identified as Jewish should therefore be

TABLE 2

ANALYSIS OF VARIANCE FOR MEMORY OF ACCURATELY IDENTIFIED PICTURES FROM SET I THAT WERE ACCURATELY IDENTIFIED AGAIN IN SET II AND SET III

Source	SS	df	MS	F
P–NP	8.11	1	8.11	3.54
Between subjects in same group	179.27	78	2.29	
Total between subjects	187.38	79		
J–NJ	119.03	1	119.03	45.61*
(J–NJ) × (P–NP)	57.59	1	57.59	22.06*
Residual	203.38	78	2.61	
Total within subjects	380.00	80		
Total	567.38	159		

* $p < .01$.

less accurate than for those identified as non-Jewish.

From a sample of 220 male subjects, the 40 high and 40 low scorers on the Levinson-Sanford anti-Semitism scale were selected, and presented with matched sets of photographs of Jewish and non-Jewish males in cap and gown from a college yearbook. Three sets of photographs were employed in a counterbalanced manner.

The findings were:

1. The high anti-Semitic subjects were both more accurate in identifying Jewish photographs, and labeled more of the photographs as Jewish than did the low anti-Semitic subjects.

2. The accuracy of the anti-Semitic subjects was not a function of the increased number of photographs identified as Jewish.

3. The anti-Semitic subjects remembered accurately significantly more Jewish photographs than non-Jewish photographs, and also were more accurate in identification of the Jewish photographs than were subjects in the low anti-Semitic group.

4. The low anti-Semitic group remembered significantly more non-Jewish photographs than Jewish, and were also more accurate in their identifications of the non-Jewish photographs than the high anti-Semitic groups.

REFERENCES

ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.

CARTER, L. F. The identification of "racial membership." *J. abnorm. soc. Psychol.*, 1948, **43**, 279–286.

ELLIOT, D. H., & WITTENBERG, B. H. Accuracy of identification of Jewish and non-Jewish photographs. *J. abnorm. soc. Psychol.*, 1955, **51**, 339–341.

LEVINSON, D. J., & SANFORD, N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339–370.

LINDZEY, G., & ROGOLSKY, S. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37–53.

SCODEL, A., & AUSTRIN, H. The perception of Jewish photographs by non-Jews and Jews. *J. abnorm. soc. Psychol.*, 1957, **54**, 278–280.

SEELEMAN, VIRGINIA. The influence of attitude upon the remembering of pictorial material. *Arch. Psychol. NY*, 1940, No. 258.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

(Received March 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 693–695

PERCEIVED CHARACTERISTICS OF ELECTED REPRESENTATIVES[1]

ROBERT R. BLAKE AND JANE SRYGLEY MOUTON
University of Texas

COMMUNICATION and problem solving by elected representatives are keys to decision making among segments of an institution where direct participation by each person is not possible. This paper deals with the question, "Are there distinguishing characteristics in the direction of greater problem solving skill as seen by their peers, among those elected to represent their group in contrast with those who are not chosen?"

The hypothesis is that the election of a member as a representative is not a random choice but is positively related to his degree of perceived social tactfulness, intellectual competence, and procedural skill. The prediction is that the higher a member's perceived social tactfulness, intellectual competence, and procedural skill, the greater the likelihood of his election to serve his group as its representative.

METHOD

Subjects

The 88 subjects were composed into 11 eight-man units, matched for educational and occupational levels, for the purpose of studying decision making in groups under autonomous conditions (Blake & Mouton, 1960). The groups engaged in a sequence of activities centering on the examination and study of personal behavior, of group action, and of intergroup problem solving. Subjects were participants in three separate human relations training laboratory programs, with three or four groups comprising a given laboratory. All subjects were from the same organization.

Sequence of Activities Leading to the Election and Interaction of Representatives

After each group had devoted approximately 12 hours to ingroup activity over a period of 3 days, it was allowed 3 hours to develop its own approach to the solution of an assigned problem that involved producing a statement of ways to improve labor-management relations. Members recognized that their solutions would be a measure of their effectiveness as a problem solving group. It was understood that solutions would be evaluated for quality through interactions of elected representatives. As is true in many representation situations, here too, group members looked forward to the selection of their solution as "best" in a win-lose sense.

Representatives were elected by rank ordering, without the opportunity of discussion or electioneering; all members from each group were evaluated according to the criterion, "Rank order the members of your group, including yourself, for serving your group as its representative."

Solutions were then dittoed and exchanged between groups. Approximately 2 additional hours were devoted to ingroup discussion of the merits of each. Members were urged to increase their understanding of the positions of the other groups by noting points of similarity and of difference and points needing clarification or elaboration.

At a meeting of the representatives, each representative then explained and clarified his group's solution in response to questions raised by representatives from other groups for an additional 1.5–2 hours. When members reported that additional interaction would not contribute to their understanding of the other group's statements, the representatives ranked the solutions from "best to worst."

Measurement of Personal Behavior of Members during Ingroup Problem Solving Activities

Another aspect of training required all group members to give one another quantitative personal feedback via judgments regarding how each person's behavior had been experienced by all the others according to the 24 different items shown in Table 1. The ratings were in private, and anonymity of raters was insured. There was no relationship between these ratings of personal behavior and the election of representatives.

Evaluations were on a nine-point scale from never, 1; hardly at all, 2; seldom, 3; a few times, 4; sometimes, 5; fairly often, 6; often, 7; almost always, 8; to always, 9. Individual scores were averaged to assign each person a rating for every item. For approximately half the groups, personal behavior ratings were made prior to the training sequence concerned with intergroup relations and for the other half following it. Since the timing of the personal evaluations reveals no significant differences in connection with assessing the perceived characteristics of representatives, the combined data are reported here.

RESULTS

The members of each group were divided into quartile divisions, according to rank received in the election of representatives. Data from all groups were combined so that each quartile division consisted of 22 subjects. Table 1 presents, for each quartile division, the mean rating received on the personal characteristic items. The three-fold categorization of items is on a priori grounds.

Results in Table 1 can be evaluated from two points of view. One is the regularity of changes from the highest to the lowest quartile division.

[1] This study was partially supported by Grant M-2447-cl, Behavior of Group Representatives under Pressure, National Institute of Mental Health.

TABLE 1

PERCEIVED CHARACTERISTICS OF INDIVIDUALS IN FOUR QUARTILE DIVISIONS FROM MOST (1st) TO LEAST (4th) PREFERRED AS REPRESENTATIVES

Type and Characteristic	Mean Rating by Quartile Division				
	4th	3rd	2nd	1st	t^a
Procedural Skill					
Takes the lead in selecting topics	3.8	5.6	6.2	6.8	6.66
Offers constructive evaluation as needed	4.8	5.7	6.8	7.1	5.34
Helps get to the meat of issues	5.1	6.2	6.8	7.1	5.26
Encourages group to high-level productivity	4.5	5.6	6.4	6.8	4.69
Provides helpful objective feedback to members	5.0	6.0	6.6	6.7	4.05
Provides good summaries when needed	4.0	5.2	6.2	6.7	3.86
Helps group stay on target	5.2	5.9	6.5	6.8	3.41
He likes to wander	3.3	3.5	3.2	2.9	1.42
Blocks the group	2.4	2.5	2.4	2.2	.80
Intellectual Competence					
Offers original ideas	4.7	5.9	6.7	7.1	6.85
Helps members express their ideas	4.4	5.4	6.2	6.6	5.41
Expresses himself clearly and concisely	5.9	6.4	7.0	7.2	3.82
Sticks to his point arbitrarily	3.1	3.5	3.5	3.3	.46
Social Tact					
Dominates and imposes his will on the group	1.9	2.6	2.9	3.5	4.10
Runs away when faced with problems	3.2	2.8	2.3	2.2	3.44
Yields to group pressures—conforms	5.4	4.7	4.4	4.0	3.41
Annoys others	3.0	4.0	4.3	4.5	2.03
Levels with other members	6.7	7.1	7.4	7.4	1.32
Effectively senses when to talk and listen	6.3	6.3	6.7	6.8	1.25
Makes others feel at ease	6.7	6.8	7.1	7.0	.81
He fights rather than works	2.3	2.6	2.5	2.6	.53
Listens with understanding to what others say	7.3	7.2	7.4	7.4	.39
He shows that he likes us	6.8	7.0	6.7	6.7	.26
Contributes without cutting others off	6.6	6.4	6.4	6.6	0

[a] Between first and fourth quartile divisions.

Progressive increases or decreases from the first to the fourth quartile division provide an indication of the stability of the relationships obtained. In addition, *t*'s have been computed between the first and fourth quartile divisions to indicate the significance of differences between extremes. Note that all items for which the *t*'s are significant also are characterized by consistent upward or downward trends from the first to the fourth quartile divisions.

The results confirm the hypothesis, clearly showing that members who are most preferred by their peers as representatives are distinguished from those who are not chosen in terms of being seen as possessing greater intellectual competence and procedural skill. They do *not* support the view that highly preferred representatives are perceievd to possess greater social tactfulness than members who are unchosen.

Procedural Skill

Seven of the nine items evaluating procedural skills show significant differences between most and least preferred representatives. In each case changes from the first quartile division to the fourth represent a steady upward trend. Five items are concerned with skill in aiding a group to function effectively at the task level, such as taking the lead in selecting topics, helping to get to the meat of issues, helping the group to stay on target, summarizing when needed, and encouraging high-level productivity. The remaining two significant items are concerned with giving constructive evaluations and providing objective feedback. Both deal with aspects of group development that help members steer themselves more effectively; that is, they involve skill in giving reactions regarding how a person feels or how others are affecting him.

The two items that fail to distinguish the extreme groups are concerned with blocking the group and with wandering off the topic. Both are actions that interfere with the group effectiveness.

Intellectual Competence

Three of the four items measuring perceived intellectual aspects show progressive increase, with differences in all cases significant far beyond the .01 level. For "offering original ideas" the difference between extremes is the most significant in the study. The other two deal with clarity in communication. The item that does not differentiate significantly is related to sticking to one's own point in an arbitrary manner, also a behavior that blocks group action.

Social Tactfulness

Only 3 of the 11 tactfulness items distinguish degrees of preference for representative beyond the .01 level, and the findings for even those three are counter to what would be expected if election as a group representative were thought of as a "popularity" contest. Though the absolute level is low for all quartile divisions, highly chosen representatives are seen as dominating and imposing their personal will on the group more than the less frequently chosen. The other two significant items can be understood in the same manner. Highly chosen persons are seen to resist group pressures toward conformity and to avoid running away when faced with problems.

Items failing to discriminate also are counter to expectation, for they are kinds of behavior that frequently are thought of as increasing an individual's personal attractiveness and therefore his prospects of being chosen as a representative. Interestingly enough, highly chosen members tend,

almost significantly, to be seen as *more* annoying than are those who are not preferred as representatives!

DISCUSSION

Much decision making within democratic institutions is through actions of persons elected by others to speak for them. Yet little is known systematically regarding the psychological processes involved in the election of a representative. We have failed to locate a single experimental study concerned directly with evaluating the personal characteristics associated with group preference of one individual over another as a representative.

The underlying theory of representational democracy *presumes* that group members elect to office those most capable of performing functions of representation. The present study confirms one aspect of this assumption by showing that persons preferred as representatives are seen as intellectually more capable and as being more skillful from a procedural point of view. The hypothesis is *not* confirmed in the area of social tactfulness as ordinarily understood. If anything opposite findings are obtained. Representatives are seen as being stronger in the sense of dominating group actions, of resisting conformity pressures, and of facing up to issues. Except for the items concerning tactfulness, the results establish the validity of the hypothesis at the perceptual level.

A logical second step now needs to be taken. The question is, "Are persons who are preferred as representatives actually more able under intergroup problem solving conditions to get their points of view effectively understood in interaction with other representatives?"

SUMMARY

The present study deals with the perceived personal characteristics of members who are preferred by their peers to serve as spokesmen in decision making interaction with representatives elected by other groups. Eighty-eight subjects composed into eight-man groups spent approximately 12 hours in autonomous ingroup activities prior to intergroup problem solving through representatives. An assigned problem served as the basis for creating differences between groups. Each group rank ordered all its members with respect to preference for a representative. Members also evaluated each other on 24 items of personal behavior categorized as measuring social tactfulness, intellectual competence, and procedural skill.

Results show clear-cut differences and are interpreted as supporting the hypothesis that members do elect as representatives those whom they see as possessing superior procedural skill and intellectual competence. They do not bear out the part of the hypothesis concerned with social tactfulness. Persons elected as representatives are *not* more outstanding than members least preferred in terms of perceived positive social—emotional characteristics such as helping, harmonizing, or making others feel liked or at ease. Rather they are seen as personally "strong" in the sense of dominating the group, resisting conformity pressures, and facing up to problems rather than running away from them.

REFERENCE

BLAKE, R. R., & MOUTON, JANE S. *Training for decision-making in groups.* New York: Putnam, 1960, in press.

(Received April 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 696-697

THE INTERACTION OF PERSONALITY AND MUSCLE TENSION IN PRODUCING AGREEMENT ON COMMONALITY OF VERBAL ASSOCIATIONS

JAMES L. KUETHE
Johns Hopkins University

A recent study by Kuethe and Eriksen (1957) related verbal stereotypy under stress to subjects' scores on the Psychasthenia (*Pt*) scale of the MMPI. Subjects with low *Pt* scores showed more stereotypy under experimentally induced anxiety or muscle tension than did similar subjects in a control group. Subjects with high *Pt* scores, however, showed less stereotypy under either stress condition than did similar subjects in a control group.

It was suggested that stereotypy represents a facilitation of response biases based on past learning. This proposition was difficult to evaluate because the subjects gave numerical responses.

The present study had two major purposes. One was to replicate the investigation of *Pt* scores and the presence or absence of tension in producing verbal stereotypy. The second was to examine the role of prior learning in verbal stereotypy.

Method

The *Pt* scale of the MMPI was administered to male students in an introductory psychology class. For the actual experiment, 40 subjects were selected at random and divided into two groups. There were 20 subjects in a "tension" group and 20 subjects in a "nontension" group. All subjects were run in individual sessions.

Each subject in the nontension group was told that the experimenter was interested in collecting data on verbal associations so that he would have a basis for comparisons in future studies that he wished to do. The subject was then read 50 stimulus words from the Kent-Rosanoff list, and the experimenter recorded the subject's association to each word.

The subjects in the tension group received the same treatment as the nontension group with two exceptions. These subjects were told that the experimenter was interested in the effect of muscle tension upon verbal associations. Each subject in this group supported a weight by pulling down on a lever while giving his associations. The weight was heavy enough to require tension in arm and shoulder muscles. Subjects indicated that the task did not require excessive effort and that they could have supported the weight for far longer.

Results and Discussion

A count was made for each of the 50 stimulus words to determine the most frequent response given by the 40 subjects. The score of an individual subject was the total number of these high frequency associations that he produced.

Within the tension and nontension groups, subjects were classified as having high *Pt* scores (*Pt*+) or low *Pt* scores (*Pt*−). The division was made at the median *Pt* score of the subjects within each group, resulting in subgroups of 10 subjects each.

Table 1 presents the mean agreement on commonality scores for the *Pt*+ and the *Pt*− subjects in the tension and nontension groups.

An analysis of variance revealed that the interaction of *Pt* scores with the presence or absence of tension was the only significant source of variance ($F = 5.45$, $p < .05$ with 1 and 36 *df*). Under tension, *Pt*− subjects show more verbal between-subject agreement than do *Pt*− subjects in a nontension situation. Tension produces slightly less verbal agreement in *Pt*+ subjects than is shown by *Pt*+ subjects in a nontension situation.

The results of the present study support the previous findings of Kuethe and Eriksen (1957). In that earlier study, two types of stereotypy were investigated. An individual could show stereotypy by not using all of the response categories available or through sequential dependencies in the generated responses.

In the present study, the mean agreement on commonality score for a group may be regarded as a measure of group response stereotypy. To the extent that a group has a high mean agreement on commonality score, the subjects in the group are giving the same responses. These responses have probably received a great deal of emphasis in the experience of individuals from similar cultural backgrounds. In response to the word "girl," a subject may say "Jane," a response that is relatively idiosyncratic, or he may say "boy," an association learned by most individuals. It is this latter type of response that seems to be facilitated by tension in subjects with low *Pt* scores.

The earlier study was concerned with variability within subjects, whereas the present study focused on variability between subjects. In both situations, however, the interaction of personality with induced tension determined the predictability of behavior in that it predicted variability. For example, *Pt*− subjects become more predictable (more stereotyped) under tension in that they do not respond randomly (Kuethe & Eriksen, 1957) and in that their responses increase in agreement on commonality.

The facilitation of strong response biases of subjects under tension would be predicted by drive theory. But this prediction does not hold for sub-

TABLE 1

MEAN AGREEMENT ON COMMONALITY FOR THE *Pt*+ AND *Pt*− SUBJECTS IN TENSION AND NONTENSION GROUPS

	Pt−	*Pt*+
Nontension Group	19.9	21.6
Tension Group	23.8	19.7

jects with high *Pt* scores. This result is interesting because the *Pt* scale correlates .92 with the Manifest Anxiety (*MA*) scale (Eriksen & Davids, 1955), and the *MA* scale is a scale for which there is a claim that it measures drive level (Taylor, 1956). Since the correlation between the two scales is of about the same magnitude as their reliability coefficient, it is probable that the instruments are for the most part measuring the same thing.

The present evidence supports the view that while the *Pt* and related "anxiety" scales may indicate initial differences in drive level between subjects, they also predict behavior in situations producing a strong drive. It is interesting that subjects with high "test anxiety" behave as though they were under relatively high drive when they are in a neutral situation but behave as though there was a reduction in drive level when they are subjected to stress. This latter phenomenon may be determined, however, by factors which override any simple relation between drive and behavior.

It may well be a mistake to estimate drive level from performance in this instance. It is possible that individuals with high scores on anxiety scales tend to use defense mechanisms when they are subjected to experimental stress that are different in kind from the mechanisms activated in low scoring subjects. For example, when subjects with high *Pt* scores are under stress, they may become vigilant. In the present study this tendency could lead *Pt*+ subjects to edit their behavior by not always giving the first response that occurred to them. Such behavior would result in lowered agreement on commonality. This would mean that the defense mechanisms of the *Pt*+ subjects would prevent the operation of any simple effect of drive upon response characteristics.

Eriksen, Kuethe, and Sullivan (1958) showed that subjects with low *Pt* scores show less awareness of their behavior than is true of high scorers. The absence of conscious interference with behavior in the present situation could permit behavior to be modified by a strong drive in the manner that drive theory would suggest.

The evidence that anxiety scales measure more than "amount of anxiety" has been discussed by De Soto and Kuethe (1959, 1960). Part of their formulation involved the possibility that a low score on an anxiety scale may often reflect a trait independent of the one indicated by a high score. Scale scores are therefore regarded by De Soto and Kuethe as a disjunctive function of amount of anxiety and the nature of the defense mechanisms activated by the testing situation.

REFERENCES

DE SOTO, C. B., & KUETHE, J. L. The set to claim undesirable symptoms in personality inventories. *J. consult. Psychol.*, 1959, **23**, 496–500.

DE SOTO, C. B., & KUETHE, J. L. On the relation between two variables. *Educ. psychol. Measmt.*, 1960, **20**, 743–749.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the Hysteria-Psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.

ERIKSEN, C. W., KUETHE, J. L., & SULLIVAN, D. F. Some personality correlates of learning without verbal awareness. *J. Pers.*, 1958, **26**, 216–228.

KUETHE, J. L., & ERIKSEN, C. W. Personality, anxiety, and muscle tension as determinants of response stereotypy. *J. abnorm. soc. Psychol.*, 1957, **54**, 400–404.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

(Received April 30, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 698-700

A FACTOR ANALYTIC INVESTIGATION OF PERSONALITY VARIABLES AND REMINISCENCE IN MOTOR LEARNING

A. W. BENDIG AND CHARLES R. EIGENBRODE
University of Pittsburgh

CONSIDERABLE interest has been aroused by the pioneer work of Eysenck (1956b) in both experimentally and theoretically relating personality variables to the phenomenon of reminiscence in motor learning. Several additional studies (Bendig & Vaughan, 1959; Ray, 1959; Rechtschaffen, 1958) have tested deductions drawn from this formulation. Briefly, Eysenck has suggested that the trait of extraversion, defined by Guilford's Rhathymia and Eysenck's (1956a) Extraversion scales, is positively correlated with the amount of reactive inhibition generated in a motor task and, consequently, with both the amount of original learning and reminiscence shown. Less surely, Eysenck also suggests that drive, as measured by Guilford's Cycloid Disposition, Taylor's Manifest Anxiety, and Eysenck's Neuroticism scales, may be positively correlated with motor learning and reminiscence.

Subsequent research has reported inconsistent results. Using measures from a reversed alphabet printing task, Rechtschaffen (1958) found no relationships between scores on the Rhathymia scale and either learning or reminiscence, and Bendig and Vaughan (1959) reported zero relationships between Extraversion and Neuroticism scale scores and motor learning. Taylor and Rechtschaffen (1959), using learning data from Rechtschaffen's original study (1958), found significant negative correlations between Manifest Anxiety scale scores and motor learning. Both Eysenck (1956b) and Ray (1959) used a pursuit-rotor task and found Extraversion scores to be negatively related to learning, but Eysenck reported the scores to be correlated with reminiscence while Ray did not. Ray also showed that Neuroticism scores were not related to either learning or reminiscence, although Eysenck reported low but statistically significant correlations among the same variables.

In considering these studies, it appears that studies using the pursuit-rotor task have demonstrated larger relationships between personality factors and learning phenomena than those using reversed alphabet printing. These studies have also restricted the investigation to one or, at the most, two personality traits, typically rhathymia or extraversion and neuroticism or manifest anxiety. Finally, previous studies have invariably dichotomized subjects on the basis of personality scale scores and tested for significant differences in mean learning or reminiscence performance rather than investigating the covariation among the measures. The present study was designed (*a*) to utilize learning and reminiscence measures from the pursuit-rotor task as described by Ray (1959), (*b*) to include a larger number of personality scale scores by using a multiscale inventory such as the Guilford-Zimmerman (1949) Temperament Survey, and (*c*) to use factor analytic procedures in investigating the relationships among these variables.

PROCEDURE

Subjects were 160 male undergraduate students enrolled in sections of a one-semester course in introductory psychology. All subjects were individually contacted in class and requested to participate. During the experimental session, the subject was first given the Guilford-Zimmerman (1949) Temperament Survey. He was then run on the pursuit-rotor with the equipment, instructions, and procedure identical with those used by Ray (1959), requiring 5 minutes of continuous practice, 15 minutes of rest, and 2 minutes of postrest practice. Each subject read or studied during the rest period. The rest period of 15 minutes was chosen to maximize reminiscence (Ray, 1959, p. 202.)

The GZTS answer sheets were scored for all 10 scales. Raw scores were used in the subsequent analysis. The variable of Motor Learning (ML) was defined as the total time on target during the 5-minute original practice period. The measure of Reminiscence (R) used was the first 20-seconds of postrest performance minus the mean of the last two 20-second prerest trials. This measure of reminiscence was chosen since it gave slightly more reliable measures of reminiscence than did the one used by Ray (first 20-second postrest trial minus the last 20-second prerest trial). A *t* test of the mean difference between the last two 20-second prerest trials did not approach significance.

The 12 variables (10 GZTS scores, ML, and R) were intercorrelated by the usual product-moment formula and six factors extracted from the matrix, using the complete centroid method. The significant factors by McNemar's criterion were rotated to oblique simple structure, using the analytic criterion developed by Pinzka and Saunders (1954).[1]

RESULTS AND DISCUSSION

The first four centroid factors were significant by McNemar's criterion. The results of the oblique

[1] The statistical analysis was accomplished using the facilities of the University of Pittsburgh Computation and Data Processing Center.

TABLE 1

CORRELATIONS OF GZTS SCALES WITH REMINISCENCE AND MOTOR LEARNING MEASURES AND OBLIQUE REFERENCE VECTOR MATRIX FOR TOTAL GROUP

(N = 160)

Variable	Correlations with		Oblique Vector Loadings				
	R	ML	Em(−)	SA	EI	Fr	h^2
General Activity	−03	−04	11	14	−03	−29	12
Restraint	−12	24**	02	−08	59	18	39
Ascendance	−02	−11	12	51	13	−19	33
Sociability	−03	−13	−08	72	−12	19	58
Emotional Stability	06	06	50	09	−01	09	27
Objectivity	12	08	54	02	−02	16	32
Friendliness	−02	14	06	−01	11	56	33
Thoughtfulness	−07	00	−10	03	44	−22	25
Personal Relations	07	13	08	29	02	50	34
Masculinity	06	11	49	−22	02	−15	31
Reminiscence (R)		−12	16	−13	−25	−13	12
Motor Learning (ML)			11	−13	33	11	15
Direction Cosines Among Reference Vectors				−58	05 05	−58 53 05	

** $p < .01$.

rotation process are given in Table 1 along with the correlations of the GZTS scales with ML and R. The Restraint scale correlated significantly with the ML measure (r = .24), but none of the other correlations were significant. The four oblique factors are the Emotionality (EM), Social Activity (SA), Extraversion-Introversion (EI), and Friendliness (Fr) factors identified in previous factor analytic investigations (Bendig, 1960a, 1960b; Bendig & Meyer, unpublished). The experimental R and ML measures load primarily on the EI factor along with the GZTS Restraint and Thoughtfulness scales. The EI factor is independent of the other factors (direction cosines of .05, .05, and .05), while the correlations among the EM, SA, and Fr factors are appreciable (cosines of −.58, −.58, and .53).

The negative loading of the R measure and the positive loading of ML on the EI factor support Eysenck's predictions regarding extraversion and reactive inhibition, assuming that the GZTS Restraint and Thoughtfulness scales are suitable indices of Eysenck's extraversion dimension. Since Eysenck considers the Rhathymia scale of Guilford's STDCR battery as the best measure of extraversion, and since the Restraint scale was derived, with reversed scoring, from the original Rhathymia scale, this assumption appears warranted. Introverts (subjects high on the Restraint scale) appear to show less reminiscence and greater motor learning than do extraverts (subjects low on the Restraint scale). If one further assumes that the GZTS Emotional Stability and Objectivity scales, like Taylor's Manifest Anxiety scale and Eysenck's MPI Neuroticism scale (Bendig, 1960b), are reversed measures of a broad "Emotionality" factor that Taylor equates with level of irrelevant drive, then the suggestion that emotional drive level is related to learning and reminiscence is unsupported. The loadings of R and ML on the EM factor are quite low, and the Emotional Stability and Objectivity scales show only low correlations with R and ML measures. Comparisons of the mean R and ML measures of subjects in the top and bottom fifth of the distributions of these two GZTS scores give no evidence of any significant differences.

Personality scale measures of extraversion appear to be related to motor learning and reminiscence, but the relationship is not very large. A major factor in the relatively low correlations is the lack of reliability of the variables, particularly the measures of reminiscence. This is evident in the correlation of the Restraint scale with R as lower than with ML, and in the smaller loading on the EI factor of the R measure. This also may be why Ray (1959) failed to find extraversion scores significantly correlated with R although his study also shows a relationship between extraversion and ML. Future research should use pursuit-rotor rather than reversed alphabet printing as the experimental task, and attention may better be concentrated on more reliable measures of motor learning rather than unreliable reminiscence indices that may confound both reactive and conditioned inhibition.

REFERENCES

BENDIG, A. W. Age differences in the interscale factor structure of the Guilford-Zimmerman Temperament Survey. *J. consult. Psychol.*, 1960, **24**, 134–138. (a)

BENDIG, A. W. Factor analysis of "anxiety" and "neuroticism" inventories. *J. consult. Psychol.*, 1960, **24**, 161–168. (b)

BENDIG, A. W., & VAUGHAN, C. J. Extraversion, neuroticism, and motor learning. *J. abnorm. soc. Psychol.*, 1959, **59**, 399–403.

EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **54**, 113–140. (a)

EYSENCK, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333. (b)

GUILFORD, J. P., & ZIMMERMAN, W. *The Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply, 1949.

PINZKA, C., & SAUNDERS, D. R. *Analytic rotation to simple structure: II. Extension to oblique solution*. Princeton, N. J.: Educational Testing Service, 1954.

RAY, O. S. Personality factors in motor learning and reminiscence. *J. abnorm. soc. Psychol.*, 1959, **59**, 199–203.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

TAYLOR, JANET A., & RECHTSCHAFFEN, A. Manifest anxiety and reversed alphabet printing. *J. abnorm. soc. Psychol.*, 1959, **58**, 221–224.

(Received May 2, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 701–704

NEW ZEALAND CHILDREN'S ATTITUDES TOWARDS MAORIS

G. M. VAUGHAN[1] AND R. H. T. THOMPSON
University of Canterbury, New Zealand

THE Maoris of New Zealand are a prominent ethnic minority numbering approximately 7% of a total population of nearly 2.5 million. However, as they have not yet achieved full integration into European culture, instances of social discrimination still occur. In the larger cities of the North Island particularly, difficulties in Maori-white relationships have been aggravated by the large postwar increase in the Maori population.

Investigations within this country of whites' attitudes towards Maoris have been few. Fitt (1955), using a Thurstone technique, found that the ethnic attitudes of students in Auckland were unfavorable during the years 1943–1947, but were relatively indifferent by 1950, a change that he related to the shift of many Maoris to the cities during the war years. An analysis carried out by Thompson (1953, 1954a, 1954b, 1955) of news items concerned with Maori affairs in New Zealand newspapers indicated that a high percentage of themes implied unfavorable attributes among Maoris. For example, Maoris are "lazy," "irresponsible," "ignorant," and "superstitious." The first use of a picture story technique to measure attitudes towards Maoris was made by Collins (1954). Thompson (1959) used a similar but more refined set of pictures to investigate the attitudes of Christchurch University students. He found some evidence of unfavorable attitudes towards Maoris, but the results were qualified by the apparent failure of some subjects to recognize certain depicted characters as Maoris.

The research reported in this paper is an exploratory investigation, using a modified Thompson picture story technique, of the development of attitudes towards Maoris among white New Zealand school children between the ages of 8 and 16 years.

Since more than 96% of Maoris live in the North Island of New Zealand, attention was also given to a possible relationship between the amount of contact with Maoris and the development of unfavorable attitudes towards them. The census figures for 1956 report that Maoris provide 4.24% of the total population (8–16 years of age) in Auckland (North Island) and 0.56% in Christchurch (South Island). Seventy-one percent of all Maori school children attend integrated schools throughout the country. From these figures, it was reckoned that of the integrated schools total population (8–16 years of age) in each city, 3.1% in Auckland (high contact, HC) and 0.4% in Christchurch (low contact, LC) were Maoris. Maori and white children attending the same school mix freely to the extent that they work together in the same classrooms, play together in the same sports teams, and mingle together without apparent discrimination during play periods.

Schools were chosen from predominantly working class areas in both cities. This restriction not only assured comparability with respect to economic status, but also ensured that contact with Maoris, who are mainly working class in New Zealand, was derived from mixing in roughly equal status groups. This control was considered essential in view of the evidence (Allport, 1954) that the degree of ethnic prejudice is related to the kind of interethnic contact involved. Allport also suggests that prejudice may be reduced by equal status contact between majority and minority groups in the pursuit of common goals.

The aims of the present research were to clarify the relationships between ethnic attitudes in New Zealand and the variables of age and contact. The hypotheses formulated were:

1. Unfavorable attitudes towards Maoris increase with age.

2. Unfavorable attitudes towards Maoris increase as the amount of contact with Maoris decreases.

METHOD

Materials and Subjects

The experimental material consisted of 10 matched pairs of TAT-type cards, depicting themes of play, school, neighbors, a friend, sport, a formal and an informal social occasion, vocation, law and order, and family. One card of each pair made up the experimental set, the other the control. Both sets of cards were photographed from originals, and they differed only in that the crucial figure in the experimental set was Maori, in the control set white. Three sample pairs of cards are shown in Figure 1.

The subjects were 120 white children, 60 boys and 60 girls, in the three age groups, 8, 12, and 16 years, selected from Auckland (HC) and Christchurch (LC) schools. These two cities were chosen as areas which differed markedly in the amount of interethnic contact available to the subjects. The HC and LC groups were matched on a residential criterion (8-years residence in a working class suburb), and a combined IQ-scholastic attainment criterion ("A" stream school classes). The 20 subjects in each city-age group were allocated to

[1] Now at Victoria University of Wellington, New Zealand.

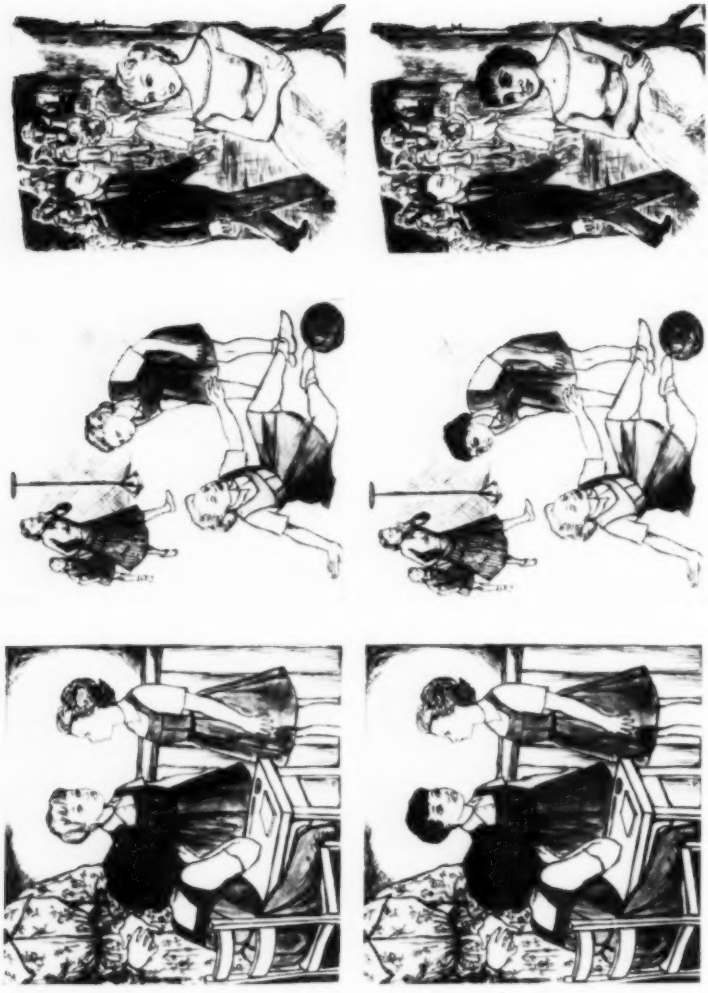


FIG. 1. Picture story test of attitudes towards Maoris. (Three sample pairs of control and experimental cards depicting themes of school, sport, and formal social occasion.)

control and experimental groups on the basis of a table of random numbers.

Procedure

Since there was considerable age range in the total sample, and therefore differences in levels of understanding test instructions, it was considered desirable to use two different sets of directions. The following instructions were given to the 8-year-olds:

This is not a test. It is a story-telling game. I am going to show you some pictures, one at a time, and I want you to make up a story about each one. These pictures are about life in New Zealand and the people in the pictures are New Zealanders. Tell me how the people in the pictures feel and think. There are no right or wrong answers. You have all the time you want.

The 12- and 16-year-olds were instructed as follows:

This is a test of creative imagination. I am going to show you some pictures, one at a time, and I want you to make up a story about each one. These pictures are about life in New Zealand and the people are New Zealanders. Tell me what has been happening in each picture, what is happening now, and then how it will turn out. Tell me how the people in the pictures feel and think. Use your imagination to the limit. There are no right or wrong answers. You have all the time you want.

Great care had to be exercised with the 8-year-old subjects to ensure adequate rapport, and sometimes it was necessary to ask them to provide a start or finish to their stories.

The scoring formula was similar to that used by Johnson (1950), who reports reliabilities ranging from 0.94 to 0.99 for this method. The stories were rated on the basis of four criteria:

1. The subject's reaction at the outset to the general setting in the picture.
2. The subject's estimate of the test figure's reaction to this situation.
3. Judgment by the subject of the test figure—e.g., whether he is esteemed or despised.
4. Outcome of the story for the test figure.

Responses were rated on a three-point scale as generally favorable, neutral, or unfavorable for each of the four criteria.

RESULTS AND DISCUSSION

The frequencies of favorable, neutral, and unfavorable responses made by control and experimental subjects at three age levels, in both HC and LC groups are shown in Table 1. Unfavorable responses indicate the degree of conflict, favorable responses the degree of cooperation recognized in the pictures by the subjects. Control subjects made their judgment when a white was the test figure, experimental subjects when a Maori was the test figure.

The results in Table 1 indicate that both HC and LC experimental groups made fewer favorable responses, more unfavorable responses, and more neutral responses than the respective control groups. However, because of considerable dependence between favorable and unfavorable measures, the statistical treatment of the data in Table 1 is limited to the percentages of unfavorable responses only made by all groups.

TABLE 1

FREQUENCY OF UNFAVORABLE, FAVORABLE, AND NEUTRAL RESPONSES TO CONTROL AND EXPERIMENTAL CARDS, IN RELATION TO AGE AND OPPORTUNITY FOR INTERETHNIC CONTACT

(Data based on 10 subjects in each cell)

Responses		High Contact Subjects		Low Contact Subjects	
	Age Level	Control	Experimental	Control	Experimental
Unfavorable[a]	8 yrs.	29	59	42	59
	12 yrs.	97	138	112	123
	16 yrs.	107	134	126	213
Favorable	8 yrs.	183	140	158	129
	12 yrs.	182	135	194	170
	16 yrs.	177	147	175	86
Neutral	8 yrs.	188	201	200	212
	12 yrs.	121	127	94	107
	16 yrs.	116	119	99	101

[a] The percentages of unfavorable responses to control and experimental cards were compared by Zubin's t test as follows:

	t	p
HC Subjects		
8 yrs.	3.53	$<.001$
12 yrs.	2.83	$<.01$
16 yrs.	2.12	$<.05$
LC Subjects		
8 yrs.	2.12	$<.05$
12 yrs.	1.41	*ns*
16 yrs.	7.78	$<.001$

An analysis of these percentages using Zubin's t test for ratio intervals (Baggaley, 1957) is appended to Table 1. This analysis shows that, in five cases out of six, both HC and LC experimental subjects gave a significantly higher percentage of unfavorable responses than did the respective control groups. In the case of the 12-year-old LC subjects, the trend was consistent. This significant increase in the degree of conflict associated with the substitution of a Maori test figure provides evidence of unfavorable attitudes towards Maoris among the subjects tested.

Age Hypothesis

HC subjects. Significant differences in the number of unfavorable responses were found between 8- and 12-year-old subjects in both control ($t = 7.07$, $p < .001$) and experimental groups ($t = 6.36$, $p < .001$). The control figures indicate a tendency

to make responses, distinctly favorable or unfavorable, to situations suggesting social conflict of any kind, while the experimental figures reflect an additional tendency to make these responses to situations specifically involving Maori-white relations. No significant differences were found between 12- and 16-year-old subjects, in either control or experimental groups, suggesting a leveling-off process in the development of social and ethnic attitudes. These results partially support, i.e., between the years of 8 and 12, the hypothesis that unfavorable attitudes towards Maoris develop as age increases.

LC subjects. Again, significant differences in the number of unfavorable responses were found between 8- and 12-year-old subjects in both control ($t = 6.36$, $p < .001$) and experimental groups ($t = 5.66$, $p < .001$), paralleling the results obtained for HC subjects. In comparing 12- and 16-year-old subjects, the difference between control groups was not significant, whereas that between experimental groups was ($t = 7.78$, $p < .001$). As with HC subjects, the control figures in this instance indicate a leveling-off process in the development of social attitudes. This process is not paralleled in the experimental figures, an aspect of the results that is amplified below.

Contact Hypothesis

No significant differences in the number of unfavorable responses were found between HC and LC control subjects at any one age level, indicating that there was little intercity difference with respect to social attitudes in general. Further, there were no significant differences between HC and LC experimental subjects at the 8- or at the 12-year-old levels. At the 16-year-old level, however, LC experimental subjects made significantly more unfavorable responses than HC experimental subjects ($t = 5.66$, $p < .001$). It is clear that, between the ages of 12 and 16 years, low interethnic contact is associated with a marked increase in unfavorable attitudes.

These results partially support, i.e., at the 16-year-old level only, the hypothesis that attitudes towards Maoris are more unfavorable among subjects experiencing minimal interethnic contact. This finding is consistent with Allport's (1954) suggestion that prejudice may be reduced by equal status contact between two groups in pursuit of common goals.

SUMMARY

A picture story technique designed to measure the attitudes of whites in New Zealand towards Maoris was employed to investigate the development of ethnic attitudes in relation to age, and contact between whites and Maoris of approximately equal status. Control and experimental sets of pictures differing only in whether the key figure was depicted as Maori or white were presented to 40 white subjects at each of the age levels 8, 12, and 16 years. Responses were scored as favorable, neutral, or unfavorable. It was concluded that:

1. Unfavorable attitudes towards Maoris increased significantly among both HC and LC subjects between the ages of 8 and 12 years.

2. Unfavorable attitudes towards Maoris increased significantly among LC subjects between the ages of 12 and 16 years.

The second conclusion supports the suggestion that prejudice may be reduced by equal status contact.

REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge: Addison-Wesley, 1954.

BAGGALEY, A. R. A table to facilitate comparison of proportions by slide rule. *Psychol. Bull.*, 1957, **54**, 250–252.

COLLINS, C. S. *A picture-story test of attitudes towards Maoris.* Unpublished MA thesis, University of Canterbury Library, 1954.

FITT, A. B. A study of racial attitudes during and after the war by the Thurstone technique. *Brit. J. Psychol.*, 1955, **46**, 306–309.

JOHNSON, G. B. An experimental projective technique for the analysis of racial attitudes. *J. educ. Psychol.*, 1950, **45**, 257–278.

THOMPSON, R. H. T. Maori affairs and the N.Z. press: I. *J. Polynes. Soc.*, 1953, **62**, 366–383.

THOMPSON, R. H. T. Maori affairs and the N.Z. press: II. *J. Polynes. Soc.*, 1954, **63**, 1–6. (a)

THOMPSON, R. H. T. Maori affairs and the N.Z. press: III. *J. Polynes. Soc.*, 1954, **63**, 216–227. (b)

THOMPSON, R. H. T. Maori affairs and the N.Z. press: IV. *J. Polynes. Soc.*, 1955, **64**, 22–34.

THOMPSON, R. H. T. European attitudes to Maoris: A projective approach. *J. Polynes. Soc.*, 1959, **68**, 205–210.

(Received May 5, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 705-708

PERCEIVED DIFFERENCES IN TASK COMPETENCE AND ASYMMETRY OF DYADIC INFLUENCE[1]

MELVYN D. CRONER AND RICHARD H. WILLIS[2]
Carnegie Institute of Technology

A number of variables have been experimentally demonstrated to effect the amount and direction of social influence in dyads. Over 20 years ago Sherif (1937) presented data suggesting that the prestige of a partner is an important factor in the amount of convergence of judgments shown by the other member of a dyad. This finding has been further documented by several more recent experiments (Bray, 1950; Kelman, 1950; Mausner, 1953; Mausner & Block, 1957; Zeaman, 1956). The importance of prior success or failure on the same task has been shown by Kelman (1950) and by Mausner (1954a). At least one previous study has investigated the role of prior success or failure of a partner on a *different, but related*, task (Mausner, 1954b). Subjects who observed their partner's success on the prior related task showed greater convergence than did those who observed their partner's failure. "Partners" were actually confederates of the experimenter, however, and it was thus not possible to make direct comparisons within dyads of the amounts of influence in each direction.

The present investigation was designed to allow such direct comparisons within dyads. It was desired to compare the amount of influence each subject exerted on the other in a judgmental situation, following the induction of perceptions of differential competence on a prior related task. The primary hypothesis may be stated formally as follows: Subjects who have been successful, on a prior related task, have more influence on their unsuccessful partners than these partners have on them.

In order to conduct a direct test of this hypothesis, an experimental situation is required in which the influence of each member of a dyad on the other can be measured separately. Customarily used tasks, such as judging the amount of apparent movement in the autokinetic situation or estimating the length of lines, are not well suited for this purpose. For example, if one subject shifts from a mean judgment of 3 inches to one of 6, while the other shifts from a mean of 10 inches to one of 6, there is some question as to which subject has been the more influenced. By comparing absolute amounts of shift, the conclusion is reached that the second subject has been influenced more, but if ratios are taken, the opposite conclusion is indicated.

A second disadvantage of such tasks is that they require a prior session, in which each subject makes judgments by himself, in order to establish a baseline from which any shift is computed.

With these considerations in mind, the second author devised a technique, to be described, that permits the simultaneous measurement of both components of dyadic influence without requiring the prior determination of individual baselines.

Two experiments are reported. The first lent no support to the primary hypothesis. A second experiment was then designed and conducted in order to test a secondary hypothesis concerning the negative results of the first experiment. The second experiment clearly supported the primary hypothesis, and, in doing so, also supported the secondary, *post hoc* hypothesis.

EXPERIMENT I

Method

Subjects and stimuli. Subjects were 20 male and 14 female undergraduates recruited from introductory psychology classes at the Carnegie Institute of Technology. Two stimulus series were employed in the subsequent task. The first consisted of 50 jars of various sizes and shapes, each containing between 250 and 1,500 beans. The second consisted of 50 lines drawn on cardboard, ranging from 3 to 15 inches.

Procedure. Subjects were run in like-sexed pairs. Prior to arrival they knew only that the experiment concerned accuracy of judgment. Upon arrival at the laboratory the experimenter informed the subjects that their first task was to take the Cardall Test of Practical Judgment. It was indicated that the Cardall test is widely used in industry and elsewhere and has been shown to be an excellent predictor of general judgmental ability. Subjects were allowed 20 minutes, after which the experimenter collected the answer sheets and pretended to grade them.

One of the subjects was randomly chosen to be accorded low perceived competence on the prior task,

[1] A paper based on the research reported was read at the 1960 American Psychological Association convention in Chicago.

[2] The authors would like to acknowledge the many helpful comments and suggestions made by Bernard Mausner of the Graduate School of Public Health, University of Pittsburgh.

while the other was accorded high perceived competence (referred to hereafter as Low and High, respectively). After the sham scoring was completed, the experimenter approached High and addressed him as follows:

> Mr. _____? Well, this is terrific! Your score was the second highest I've had yet and is one of the best possible on this test. Very good!

The experimenter addressed Low as follows:

> Mr. _____, are you sure you understood the directions I gave you for taking this test? [Low always responded that he did.] Well then, are you feeling all right today? Perhaps you're sick? [Low always answered that this was not the case.] Well, Mr. _____, you score is the worst I've had yet, and is almost as low as one can score. This experiment is not a joke, and if you tried to do poorly, or didn't try at all, say so now and you may leave. [Low always indicated that he had tried to do his best.] Mr. _____, if this is the best you can do, I would strongly suggest that you make an appointment at the Bureau of Measurement and Guidance to talk with one of the counselors. At any rate, let's get on with the rest of the experiment and see how you do there.

The experimenter made his remarks to each subject loud enough to be clearly overheard by the other.

After the perceived differences in related task competence were established, subjects proceeded to the subsequent task. Each stimulus was clearly visible to both subjects, and the experimenter announced a standard against which the stimulus was to be compared. Subjects were required to judge aloud whether the stimulus was longer or shorter, or more or less numerous, than the standard. The experimenter indicated that stimuli would exceed the standard half the time and would be exceeded by the standard half the time. In actuality, the standard was in every case exactly equal to the stimulus magnitude. High went first on the odd numbered trials, while Low went first on the even numbered trials.

Computation of scores. Data from odd numbered trials and even numbered trials were treated separately. The expected proportion of trials that both subjects would respond "plus" (stimulus larger than standard) was determined by multiplying the proportion of times High responded "plus" by the proportion of times Low responded "plus." A corresponding product was formed for the minus responses. The sum of these two products was subtracted from the actual proportion of agreements. If this difference was positive, the subject going second tended to imitate his partner, while a negative difference indicated an opposition tendency. An imitation score was computed for each subject by algebraically summing his scores on the two stimulus series (beans and lines), and an imitation score was computed for each dyad by algebraically summing the imitation scores of the two subjects. Difference scores were computed for each dyad by subtracting the imitation score of High from that of Low.

Results

The results of Experiment I are presented in Table 1. Because sex differences were negligible, only means for all dyads are given. The amount of imitation behavior exhibited by subjects was highly significant, with all five values of t exceeding the .0001 level. On the other hand, the degree to which High imitated Low was not significantly different from the degree to which Low imitated High as indicated by the t test for correlated means. We conclude that the hypothesis under investigation received no support from the data.

TABLE 1
AMOUNT AND DIRECTION OF IMITATION: EXPERIMENT I

	M	SE	t	df	p
Imitation Scores					
Lines	.2786	.0408	6.83	16	<.0001
Beans	.3180	.0404	7.87	16	<.0001
High subjects	.3155	.0334	9.45	16	<.0001
Low subjects	.2811	.0276	10.18	16	<.0001
Total	.5966	.0519	11.50	16	<.0001
Asymmetry of Imitation Scores					
Lines	−.0004	.0202	0.02	16	.50
Beans	−.0340	.0318	1.07	16	>.15
Total	−.0344	.0493	0.70	16	>.25

EXPERIMENT II

One possible explanation for the lack of significant asymmetry in Experiment I is that the experimental manipulation was not successful. However, the highly visible reactions of subjects made this seem quite improbable. A more likely explanation was that subjects failed to perceive the abilities tapped by the Cardall test as having any appreciable similarity to the bean and line judging. Consequently, a decision was made to run a second experiment, employing a prior task that would be perceived by subjects as having greater similarity to the subsequent task.

Method

Subjects and stimuli. Subjects were 20 male undergraduates recruited from introductory psychology classes. The same stimuli were used for the subsequent task as had been used in Experiment I.

Procedure. Exactly the same procedure was followed as in Experiment I, except that the Cardall test was replaced by a series of 36 irregularly drawn areas of various sizes. Prior to judging these areas subjects listened to a rather elaborate explanation by the experimenter which was designed to bring out the similarity between the area judging on the one hand and

TABLE 2
AMOUNT AND DIRECTION OF IMITATION: EXPERIMENT II

	M	*SE*	*t*	*df*	*p*
Imitation Scores					
Lines	.2791	.0503	5.55	9	.0003
Beans	.3273	.0537	6.10	9	<.0001
High subjects	.1587	.0376	4.22	9	.0012 > *p* > .001
Low subjects	.4477	.0564	7.94	9	<.0001
Total	.6064	.0672	9.02	9	<.0001
Asymmetry of Imitation Scores					
Lines	.1780	.0348	5.12	9	.0005
Beans	.1110	.0453	2.45	9	<.02
Total	.2890	.0686	4.21	9	.0012 > *p* > .001

the bean and line judging on the other. As before, the experimenter announced a correct standard, but subjects were required to respond "plus" or "minus." Judgments of the areas were recorded on answer sheets, so that neither subject knew the responses given by his partner. In "grading" the answer sheets and in inducing differential perceptions, the experimenter followed the procedure used in Experiment I.

Results

The results of Experiment II are presented in Table 2. As with Experiment I, the extent to which subjects imitated their partners was highly significant. In contrast to Experiment I, however, Low subjects imitated High subjects to a significantly greater degree than conversely, as indicated by the *t* test for correlated means. The overall difference on all stimuli closely approached the .001 level. Furthermore, the mean imitation score for Low subjects was almost three times as large as the mean for High subjects. We may therefore accept the hypothesis under investigation with confidence.

Table 3 gives the percentage of total influence that consisted of imitation, for both experiments. Total amount of influence was computed by summing the absolute magnitudes of the scores earned by each subject on each stimulus series, whereas amount of imitation was determined, as previously stated, by computing the corresponding algebraic sum. The difference between the two is due to the presence of negative scores. A negative score indicates that the subject agreed with his partner less often than would be expected by chance.

The percentages of imitation for Experiment I are so large as to make differences between stimulus series and partners negligible. The percentages for Experiment II are also large, although not so large as in the first experiment. The rather appreciable difference between the percentage for High and that for Low is another manifestation of the asymmetry between partners. Not only do High subjects imitate their partners less in an absolute sense, but they also exhibit a smaller degree of imitation relative to the total amount they deviated from the chance level of agreement. In other words, High subjects showed more pronounced opposition tendencies than did Low subjects.

TABLE 3
PERCENTAGE OF TOTAL INFLUENCE ATTRIBUTABLE TO IMITATION

	Experiment I	Experiment II
Lines	97.7	87.2
Beans	97.4	95.7
High subjects	99.3	74.1
Low subjects	95.5	100.0
Total	97.5	91.6

DISCUSSION

The almost perfect symmetry found in Experiment I, coupled with the highly significant degree of asymmetry found in Experiment II, indicates that the gradient of generalization for influence from one task to another is relatively steep. In the absence of a set of tasks scaled on perceived similarity, it is not possible to determine the exact shape of this generalization curve, but there seems little doubt that the outcome of a prior task will have little or no effect during the subsequent task unless the two tasks are perceived as being highly similar.

Some supplementary support for this conclusion comes from the extremely high degree of significance reached by the tests for the amount of imitation in Experiment I. There is every reason to believe that a technique so highly sensitive to the presence of influence would also be sensitive to any directional difference that might be present.

The conclusion of a relatively steep gradient of generalization does not rule out the possibility of a highly general evaluation of one individual by another. Subjects in the present experiment were previously unacquainted, and as a result each subject had only the announced outcomes on the prior task to guide him in making relative evaluations. If *several* prior tasks had been employed, and if one subject were always seen to do much better than the other, it is quite possible that highly generalized relative evaluations would result. Certainly in cases of extended interpersonal con-

tact outside the laboratory such generalized evaluations occur at times, and the question of the necessary conditions warrants systematic investigation.

The relative amount of opposition behavior was greatest among High subjects of Experiment II. Although the difference between High subjects in the two experiments only approached significance ($.10 > p > .05$), it raises some interesting possibilities for future investigations. One might hypothesize that, other things being equal, status differentials in a group lead to a lower overall level of group conformity. In a stratified overall group, the high status groups become positive reference groups for the members of the low status groups. The low status groups, however, take on the properties of negative reference groups for high status members. If and when the low status groups learn to imitate successfully the high status groups, these high status groups should become motivated to change their patterns of behavior. In a more egalitarian group, pressures towards imitation and conformity are given free rein due to the absence of well developed negative reference groups. Consider the hypothetical case of complete absence of status differences. No member or group could utilize superior status as a basis for effecting an innovation. If such an attempt were made, the remaining members would have no motivation to follow. One would predict that a state of equilibrium would be approached at which a common pattern of behavior would be adopted by all members of the group.

SUMMARY

The problem investigated was that of the effect of perceived task similarity and perceived differences in task competence on the asymmetry of influence between members of dyads.

Two experiments were conducted. In both experiments a *prior* task, in which perceptions of differential task competence were induced, was followed by the *subsequent* task of judging the magnitudes of 100 stimuli. In Experiment I, the induction of perceptions of differential prior task competence failed to produce a significant degree of asymmetry of influence during the subsequent task. It was hypothesized that the lack of significant asymmetry was due to the failure of subjects to perceive the prior task as requiring skills similar to those required by the subsequent task.

With the above *post hoc* hypothesis in mind, a second experiment was designed and conducted. The prior task was replaced by one of greater similarity to the subsequent task, and instructions were rewritten in order to emphasize this similarity. Experiment II yielded a highly significant degree of asymmetry of influence. It is concluded that perceived task competence is an important determinant of the amount and direction of social influence, but that its effectiveness depends in a crucial way on perceived task similarity.

REFERENCES

BRAY, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 64–84.

KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

MAUSNER, B. Studies in social interaction: III. The effect of variation in one partner's prestige on the interaction of observer pairs. *J. appl. Psychol.*, 1953, **37**, 391–393.

MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 65–68. (a)

MAUSNER, B. The effect of one partner's success or failure in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557–560. (b)

MAUSNER, B., & BLOCH, BARBARA, L. A study of the additivity of variables affecting social interaction. *J. abnorm. soc. Psychol.*, 1957, **54**, 250–256.

SHERIF, M. An experimental approach to the study of attitudes. *Sociometry*, 1937, **1**, 90–98.

ZEAMAN, D. The account of the demonstration contributed by Dr. David Zeaman. Cited and quoted in M. Sherif & Carolyn W. Sherif (Eds.), *An outline of social psychology*. (Rev. ed.) New York: Harper, 1956. Pp. 610–612.

(Received May 9, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 709-712

SUCCESSIVE RESPONSES TO SIMULTANEOUS STIMULATION IN ELDERLY PATIENTS WITH MEMORY DISORDER[1]

JAMES INGLIS AND R. E. SANDERSON
Queen's University, Kingston, Ontario

CHANGES in learning ability with age can be shown to affect the performance of otherwise normal people (Welford, 1958), and severe disturbances of the learning processes constitute an important part of that kind of disorder of the aged commonly called "senile dementia" (Inglis, 1958).

Welford (1956) has argued that the key to the understanding of the crucial nature of these and other performance decrements with age may lie in a better knowledge of the processes underlying short-term retention. It has been shown, however, that on the orthodox digit span test, often used clinically as a measure of short-term retention, performance falls off very little with age (Bromley, 1958) and does *not* appear to discriminate well between elderly psychiatric patients with and without memory disorder (Inglis, 1957).

A number of experiments carried out by Broadbent (1958) and his associates with young adult subjects have shown that certain modifications in the way of presenting digit span stimuli may serve to elucidate some of the problems raised by performance on this apparently simple task. Broadbent has suggested that whereas the orthodox way of presenting the digit span stimuli may involve mainly a simple perceptual process, the simultaneous presentation of different parts of a span through different sensory channels (e.g., both ears at once) may require, in addition, the participation of a simple storage system, which itself may provide the basis for short-term memory.

Thus, in one experiment Broadbent (1957) showed that when digit span stimuli were relayed at a speed of 2 per second through headphones to each subject, one half of the span to one ear and, simultaneously, the other half of the span to the other ear, then the subjects tested under these conditions could reproduce the digits *sequentially*. Furthermore, in such reproduction, elements from one half of the span were never alternated with elements from the other half. The first half-set recalled, however, commonly contained fewer errors than the second half; this constituted a kind of serial-order effect. Broadbent (1956), in an experiment in which the half-spans were presented to eye and ear simultaneously, obtained similar results, demonstrating that these phenomena are not peculiar to hearing but may be generalized to different sense modalities.

Broadbent suggested two kinds of mechanism which may underlie such performances on the digit span test. First, there is a "p-system," which can only pass information successively. Second, there is an "s-system," which can store *excess* information, arriving, for example, when the p-mechanism is already fully occupied in transmitting information from another channel. Broadbent (1957) tried to find out experimentally if the s-system could be regarded as operative even in the orthodox digit span test or if it could best be regarded as a kind of emergency device, only called upon when excess information must be handled. In this case, with dichotic stimulation, the *order* of recall of the two half-spans was prescribed by the experimenter. The recall order was prescribed to some of the subjects *before* the stimuli were presented and to some of the subjects after presentation of the stimuli.

If it were the case that all the stimuli passed through the s-system before entering the p-system, then the serial-order effect would be expected to appear, regardless of whether the recall order had been prescribed before or after the stimuli were presented. If, on the other hand, only part of the information passed through the s-system, the other part going straight through the p-system, then the serial-order effect would *not* be expected to appear in the situation in which the recall order had been prescribed *after* the presentation of the stimuli.

This experiment did not produce completely unequivocal results. In the situation in which recall order was imposed before stimulation, the serial-order effect did appear. In the situation in which recall order was specified after stimulation, the serial-order effect only appeared in the early stages of practice. In the later stages, this effect almost disappeared, the two half-sets of digits being recalled with about equal efficiency. This suggests that while initially all the data may go through the s-system before being passed through the p-system, in the later stages of practice only part of the data passes through the s-system first. Broadbent admits that this outcome is not fully conclusive. Nevertheless, he maintains that it is

[1] This study is part of a project carried out with the assistance of a Canadian National Health Grants Administration Mental Health Grant (Project No. 605-5-285). The authors are also indebted to the Superintendent and staff of the Ontario Hospital, Kingston, for their help and cooperation.

most likely that, in the ordinary immediate memory test, the information probably enters the p-system at once without passing first through the s-system.

This notion is supported by other evidence (Inglis, 1957) which shows that elderly psychiatric patients with memory disorder (and hence showing presumptive evidence of some breakdown of their storage processes) may do about as well on ordinary digit span tests as patients of the same age but without memory disorder. Zangwill (1943) had, of course, previously pointed out that restriction of immediate memory as measured by the digit span test may be regarded as the exception, even in cases showing a severe retention defect.

If memory disorder in elderly psychiatric patients is, in part at least, based upon some breakdown of the kind of storage mechanism postulated by Broadbent, then we may expect such patients to show specifiable disturbances of recall in the dichotic stimulation situation relative to the performance of a matched group of patients without memory disorder. The experiment reported here was designed to investigate this notion.

METHOD

Subjects

The experimental group comprised 15 elderly patients clinically estimated to be suffering from undoubted memory disorder. This estimate was made on the basis of a complaint about the patient's memory made by the doctor in charge of the case, by the ward staff, or both.

A control group of 15 patients without memory disorder was also tested. The age range of the experimental group was 68–83; that of the control group was 66–84. The Wechsler Verbal scale IQ range in the experimental group was 77–103; in the control group it was 75–112. There were 4 men and 11 women in each group, all psychiatric patients of the Ontario Hospital, Kingston, Ontario. They were almost all of the rural working class. The women were, for example, almost all farmer's and farm worker's wives. All the patients (both experimental and control) were cooperative during testing, and none of them suffered any sensory or motor deficit so severe as to interfere with test performance. The experimental group was mainly composed of cases diagnosed as senile dementia, usually having been admitted to hospital for difficulties associated with failing memory. The control group was principally composed of functionally disordered patients diagnosed as cases of depression and paranoid schizophrenia.

Tests

The WAIS (Wechsler, 1955) was used for matching the two groups on verbal ability (Verbal IQ) and on average digit span forward. This test was also used to check on a previously observed discrepancy between the Verbal and Performance scales (Inglis, 1959) in a memory-disordered group.

TABLE 1

DIGITS USED FOR DOUBLE STIMULATION TEST

Channel 1	Channel 2
Practice Series	
8	Blank
Blank	5
8	5
Test Series	
6	2
9	7
7	4
5	8
32	94
18	65
37	49
57	86
658	417
432	869
571	243
769	354
6235	1943
4863	2547
2719	8475
9327	6351

The apparatus used to administer the binaural stimuli consisted of a Webcor stereophonic tape recorder (Model EP2821-5B) played into a pair of Edison (Model 30) earphones. Different sets of digits were recorded on each channel as shown in Table 1. The digits were presented at the rate of 1 per second.

Each subject was first told, "You are going to hear a number. Tell me what you hear." The spoken digit 8 was then played on Channel 1. If the subject responded correctly, then the same procedure was repeated for the digit 5 on Channel 2. If the subject failed to respond or gave the wrong number, the volume was increased until the correct response was made.

Second, the subject was told, "Now you are going to hear two numbers together, one in each ear. Tell me what you hear." The two channels then played the spoken digits 5 and 8 simultaneously. If the subject responded with the correct digits (i.e., 5-8 or 8-5), then the experimenter went on to the test series. If not, this item of the practice series was repeated once more.

The foregoing constituted a practice series which served to ensure that any difference found between the groups could not merely be due, say, to sensory differences. It also got the subjects used to the situation.

Third, each subject was told, "Now you are going to hear n numbers, n/2 in each ear" (where n was 2 or 4 or 6 or 8). "Tell me what numbers you hear." These instructions were repeated at the beginning of each series of a new length. Between each of the items within a series, the experimenter said, "Now what numbers do you hear?"

This series of tests were scored as follows: The first correct digit repeated in each series determined which channel was taken as the "dominant" one for that item

TABLE 2

MEANS, STANDARD DEVIATIONS, AND THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE MEANS OF THE CRITERION GROUPS ON THE VARIABLES EXAMINED

Variable	Experimental Group		Control Group		Between Group Differences	
	M	σ	*M*	σ	*t*	*p*
Age	75.73	4.42	75.00	4.46	0.57	*ns*
WAIS IQs Verbal scale	89.00	9.27	95.40	10.38	1.78	*ns*
Performance scale	72.07	6.43	88.53	11.02	4.99	<.01
Verbal minus Performance	16.93	8.29	6.87	8.36	3.27	<.01
Full Scale	80.67	7.57	92.07	10.39	3.44	<.01
Digit Span Forward	5.93	1.44	6.53	0.83	1.36	*ns*
Dichotic stimulation						
		p within group difference		*p* within group difference		
1 Digit						
Dominant	0.98	0.14 } <.01	1.00	0.00 } *ns*	0.55	*ns*
Nondominant	0.47	0.41 }	0.95	0.44 }	3.10	<.01
2 Digits						
Dominant	1.83	0.20 } <.001	1.95	0.20 } *ns*	1.64	*ns*
Nondominant	0.09	0.63 }	1.80	0.45 }	8.30	<.001
3 Digits						
Dominant	2.42	0.48 } <.001	2.90	0.17 } <.01	3.64	<.01
Nondominant	0.05	0.14 }	2.15	0.67 }	11.86	<.001
4 Digits						
Dominant	2.82	1.06 } <.001	3.63	0.45 } <.001	2.72	<.02
Nondominant	0.09	0.63 }	0.95	0.66 }	3.55	<.01

by that subject; the second half of the total digit span was called "nondominant." The score obtained was the average number of correct responses for each half digit set taking, of course, digit position in the series into account.

RESULTS AND DISCUSSION

The means and standard deviations and the significance of the difference between the means of the experimental and control groups are shown in Table 2. The significance of the differences was estimated by the uncorrelated *t* test (McNemar, 1949).

Previous findings are confirmed by the significance of the Verbal Performance difference on the WAIS.

The differences between the groups in the dichotic stimulation experiment are significant in the "nondominant" channel for all the lengths of digit series used. This suggests that the clinical picture of memory disorder in elderly psychiatric patients is based, at least in part, upon a breakdown of the kind of storage mechanism (s-system) proposed by Broadbent to account for the capacity of young normal adults to respond successively to information delivered to them simultaneously.

It can also be seen that significant differences of lesser magnitude appear between the groups in their responses to information in the dominant channel for the two longest half-sets (i.e., 3 and 4 digits). No explanation can at present be offered for this finding.

Some of the differences between the accuracy of recall in the dominant and the nondominant channels *within* the experimental and control groups are also statistically significant. This finding is in line with the notion of a kind of serial-order effect in the recall of the second half of each digit series. It is also congruent with the present hypothesis that this effect should appear even in the shortest series in the case of the experimental group, although it does not appear in the control group until a total digit series length of six figures is reached.

It may even be possible to apply the results obtained in this investigation to secure a clinical test of memory disorder which would be just as convenient to use as the orthodox digit span but much more likely to be valid.

SUMMARY

This study was designed to test the notion that memory disorder, clinically ascertained in elderly persons, may be related to a breakdown in the kind of storage system which has been adduced to explain successive responses to simultaneous stimulation in normal young adults.

Two groups of 15 elderly psychiatric patients, one group of patients with memory disorder and one group without, were tested. The results obtained support the hypothesis put forward.

REFERENCES

BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, 1956, **8**, 145–152.

BROADBENT, D. E. Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, 1957, **9**, 1–11.

BROADBENT, D. E. *Perception and communication.* London: Pergamon, 1958.

BROMLEY, D. B. Some effects of age on short term learning and remembering. *J. Geront.*, 1958, **13**, 398–406.

INGLIS, J. An experimental study of learning and "memory function" in elderly psychiatric patients. *J. ment. Sci.*, 1957, **103**, 796–803.

INGLIS, J. Psychological investigations of cognitive deficit in elderly psychiatric patients. *Psychol. Bull.*, 1958, **55**, 197–214.

INGLIS, J. Learning, retention, and conceptual usage in elderly patients with memory disorder. *J. abnorm. soc. Psychol.*, 1959, **59**, 210–215.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

WECHSLER, D. *Wechsler Adult Intelligence Scale Manual.* New York: Psychological Corporation, 1955.

WELFORD, A. T. Age and learning: Theory and needed research. In *Symposium on experimental gerontology.* (Suppl. 4) Basel: Birkhauser, 1956.

WELFORD, A. T. *Ageing and human skill.* London: Oxford Univer. Press, 1958.

ZANGWILL, O. L. Clinical tests of memory impairment. *Proc. Roy. Soc. Med.*, 1943, **36**, 576–580.

(Received May 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 713-715

INTERPERSONAL ATTRACTION AND ATTITUDE SIMILARITY[1]

DONN BYRNE[2]
University of Texas

In investigating the direction and the strength of the affect engendered between the two participants in a dyad, we may arrange the expressed feelings of each individual along a continuum ranging from strongly positive to strongly negative. The accurate prediction of interpersonal attraction and repulsion in such relationships will undoubtedly require that we secure knowledge about several classes of independent variables.

Probably the most obvious and also best documented variable is that of propinquity. Studies in a wide variety of settings have shown that physical and functional distance influence interaction and interpersonal attraction (Byrne, in press). Once the environmental situation permits or encourages interaction, affiliation need should be helpful in predicting individual differences in interpersonal behavior (Atkinson, Heyns, & Veroff, 1954; Schachter, 1959). A third class of variables consists of the overt stimulus properties of each individual to which other individuals would be expected to respond on the basis of generalization from previous interpersonal interactions.

Once interaction has begun, reciprocal reward and punishment is proposed as the crucial determining factor. It has been suggested (Newcomb, 1956) that attraction between persons is a function of the extent to which reciprocal rewards are present in their interaction; perhaps dislike is a function of reciprocal punishments. A special subclass of this variable would be perceived similarity and dissimilarity of the attitudes of two individuals. It can be assumed that persons in our culture have well established learned drives to be logical and to make a correct report of the environment. Those who seem deficient in this respect are generally categorized as being uninformed, of low intelligence, immoral, and/or as being out of contact with reality. It is primarily through consensual validation that we determine whether we or anyone else is logical or correct in interpreting environmental events. Hence, any time that another person offers us validation by indicating that his percepts and concepts are congruent with ours, it constitutes a rewarding interaction and, hence, one element in forming a positive relationship. Any time that another person indicates dissimilarity between our two notions, it constitutes a punishing interaction and thus one element in forming a negative relationship. Disagreement raises the unpleasant possibility that we are to some degree stupid, uninformed, immoral, or insane. An alternative possibility is that it is the other person who is deficient in one or more of these characteristics. Probably other variables, such as the importance of the issue to each individual, contribute to the effect.

A number of studies have found greater similarity among friends than among nonfriends with respect to a variety of issues (Bonney, 1946; Loomis, 1946; Newcomb, 1956; Precker, 1952; Richardson, 1940; Winslow, 1937). A few studies of a more experimental nature also support the notion of a relationship between attitude similarity and interpersonal attraction (Jones & Daugherty, 1959; Smith, 1957).

In order to test the proposition that the effect of attitude similarity is a causative one and to test some implications arising from the preceding speculations about the reason for the effect, it was hypothesized that (*a*) a stranger who is known to have attitudes similar to those of the subject is better liked than a stranger with attitudes dissimilar to those of the subject, (*b*) a stranger who is known to have attitudes similar to those of the subject is judged to be more intelligent, better informed, more moral, and better adjusted than a stranger with attitudes dissimilar to those of the subject, and (*c*) a stranger who is known to have similar attitudes on issues important to the subject and dissimilar attitudes on unimportant issues is better liked and is evaluated more positively on the other four variables than a stranger for whom the reverse is true.

METHOD

Attitude measure.[3] On the basis of a pilot study, 26 issues were selected for inclusion in an attitude and

[1] A portion of this paper was read at the meetings of the Southwestern Psychological Association, Galveston, 1960.

This investigation was supported in part by a research grant (EF-140) from the University of Texas Research Institute.

[2] From the Laboratory for Personality Research.

[3] The attitude measure, response frequencies, and the rating scales for interpersonal attraction and evaluation have been deposited with the American Documentation Institute. Order Document No. 6770 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

opinion scale. Each issue was presented in a seven-point scale. The issues ranged from those thought to be extremely important by the pilot subjects (e.g., integration, God, premarital sex relations) to those considered to be of minor importance (e.g., western movies and television programs, classical music, politics).

Procedure. The attitude scale was administered to 64 students (36 male, 28 female) enrolled in an introductory psychology course at the University of Texas. Response heterogeneity differed from item to item, but there was moderately wide diversity of opinion among the 64 subjects. After filling out the attitude scale, the subjects were asked to indicate which they believed to be the 13 most important and 13 least important issues.

Two weeks later they were falsely informed that the attitude scale had been given as part of a study in interpersonal prediction. They were told that individuals in another class had been given the same scale that they took, students in the two classes were matched on the basis of sex, and they were to be given each other's tests (name removed) in order to determine how much they could learn about one another from this information alone.

Actually the questionnaire they received at this time was a fake one made up by the experimenter. The subjects had been randomly divided into four groups; one group received attitude scales filled out exactly the same as theirs had been, one received scales with exactly opposite views expressed, one received scales with similar opinions on the most important issues and dissimilar on the least important, and the fourth received scales with similar opinions on the least important issues and dissimilar opinions on the most important. The four groups of subjects did not differ significantly in their initial responses to any of the 26 issues.

Interpersonal attraction and evaluation. A rating scale was used as the measure of interpersonal attraction and evaluation with each dependent variable represented in a seven-point scale. As a measure of interpersonal attraction, subjects were asked to indicate how well they felt they would like this person and whether they believed they would enjoy working with him (or her) as a partner in an experiment. Four scales dealt with evaluation; the subjects were asked for their judgments as to the other student's intelligence, knowledge of current events, morality, and adjustment.

RESULTS

First Hypothesis. Table 1 shows the comparisons of the two groups on each of the dependent variables. The first hypothesis was overwhelmingly confirmed for each of the two attraction scales. The group with attitude scales filled out the same as their own (SA) indicated significantly more positive feelings toward the "stranger" than did the group which received scales indicating dissimilar attitudes (DA). Each difference was significant at less than the .001 level.

Second Hypothesis. As is indicated in Table 1, the second hypothesis was also confirmed. The SA group rated the "stranger" significantly higher than did the DA group on intelligence, knowledge of current events, morality, and adjustment. Again, each difference reached a level of significance beyond the .001 level.

Third Hypothesis. The third hypothesis, concerning the influence of important vs. unimportant

TABLE 1
COMPARISON OF THE SIMILAR ATTITUDE (SA) AND DISSIMILAR ATTITUDE (DA) GROUPS ON INTERPERSONAL ATTRACTION AND EVALUATION

	SA (N = 17)		DA (N = 17)		D	t	df	p
	M	SD	M	SD				
Personal Feelings	6.53	.50	1.76	.73	4.77	21.46	32	<.001
Desirability as Work Partner	6.47	.50	2.65	1.88	3.82	7.88	32	<.001
Intelligence	5.65	.68	3.06	.87	2.59	9.37	32	<.001
Knowledge of Current Events	4.65	1.14	2.65	.91	2.00	5.51	32	<.001
Morality	5.76	.73	3.47	2.09	2.29	4.14	32	<.001
Adjustment	6.00	.84	2.71	1.15	3.29	9.36	32	<.001

TABLE 2
COMPARISON OF THE SIMILAR ON IMPORTANT ATTITUDES (SIA) AND SIMILAR ON UNIMPORTANT ATTITUDES (SUA) GROUPS ON INTERPERSONAL ATTRACTION AND EVALUATION

	SIA (N = 15)		SUA (N = 15)		D	t	df	p
	M	SD	M	SD				
Personal Feelings	4.20	1.51	2.60	1.20	1.60	3.10	28	<.01
Desirability as Work Partner	4.27	1.44	3.33	1.40	.94	1.76	28	*ns*
Intelligence	4.13	.62	3.73	1.34	.40	1.01	28	*ns*
Knowledge of Current Events	3.60	.95	3.53	.96	.07	.19	28	*ns*
Morality	5.33	1.25	3.33	1.66	2.00	3.60	28	<.01
Adjustment	4.07	1.57	2.93	1.18	1.14	2.17	28	<.05

issues, was only partially confirmed. As is shown in Table 2, the Similar on Important Attitudes Group (SIA) rated the "stranger" significantly more positively than did the Similar on Unimportant Attitudes Group (SUA) with respect to their personal feelings about him, his morality, and his adjustment. On the other three variables, the two groups did not differ.

DISCUSSION

The experimental confirmation of the first two hypotheses is very encouraging for further research designed to investigate other aspects of the relationship between interpersonal attraction and attitude similarity. It should be possible now to study the effect of attitude differences less extreme than those in the present study and to combine this variable with the others that influence interpersonal attraction in order to determine interaction effects.

Because of the fact that this group of subjects showed a degree of homogeneity of opinion on some of the attitude items, a possible alternative interpretation is that they were responding negatively to unusual and deviant beliefs rather than to disagreement per se. On 19 of the 26 issues it was possible for a subject to fall in the deviant one fourth of the group by expressing a positive or a negative opinion. The range among the subjects was from no deviant attitudes to nine; they were divided into high and low subgroups on the basis of this score. Since these "conforming" and "deviant" subgroups did not differ from one another in responding to strangers with similar vs. different attitudes, there is no evidence to support this other interpretation of the results.

The partial failure of the third hypothesis led to a comparison of all four groups on each of the dependent variables. The results suggest that the Personal Feelings scale is the most sensitive measure of interpersonal attraction. With the other five interpersonal judgment scales, additional factors apparently contribute to the variance.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

BONNEY, M. E. A sociometric study of the relationship of some factors to mutual friendships on the elementary, secondary, and college levels. *Sociometry*, 1946, **9**, 21–47.

BYRNE, D. The influence of propinquity and opportunities for interaction on classroom relationships. *Hum. Relat.*, in press.

JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *J. abnorm. soc. Psychol.*, 1959, **59**, 340–349.

LOOMIS, C. P. Political and occupational cleavages in a Hanoverian village, Germany: A sociometric study. *Sociometry*, 1946, **9**, 316–333.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

PRECKER, J. A. Similarity of valuings as a factor in selection. *J. abnorm. soc. Psychol.*, 1952, **47**, 406–414.

RICHARDSON, HELEN M. Community of values as a factor in friendships of college and adult women. *J. soc. Psychol.*, 1940, **11**, 303–312.

SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford Univer. Press, 1959.

SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. Psychol.*, 1957, **43**, 251–260.

WINSLOW, C. N. A study of the extent of agreement between friends' opinions and their ability to estimate the opinions of each other. *J. soc. Psychol.*, 1937, **8**, 433–442.

(Received May 25, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 716-719

ANXIETY, BIRTH ORDER, AND SUSCEPTIBILITY TO SOCIAL INFLUENCE[1]

FREDERICK R. STAPLES AND RICHARD H. WALTERS
University of Toronto

SCHACHTER (1959) claims that first-born children are subjected to more inconsistent nurturance than are later-born children and consequently show more dependency behavior in the form of affiliative responses. Schachter finds support for his hypothesis in studies reported by Rosonow and Whyte (1931), Sletto (1934), Sears (1950), Sears, Maccoby, and Levin (1957), and Haeberle (1958). In addition he reports experiments carried out at Minnesota which show that first-born subjects more readily change their opinions when fictitious group norms are provided (Ehrlich, 1958) and experience greater anxiety reduction through affiliation with others (Wrightsman, 1959) than do later-born subjects. The experiment reported in this paper is an attempt to test Schachter's hypothesis, using suggestibility in the autokinetic situation as an index of susceptibility to social influence.

The following hypotheses were put forward:

1. In the autokinetic situation, even when an anxiety inducing situation is not created, first-born subjects show greater suggestibility than do later-born subjects.

2. The introduction of anxiety increases susceptibility to social influence. This effect is more marked for first-born than for later-born subjects.

METHOD

Subjects

Sixty-four female undergraduates served as subjects. Thirty-two subjects were first-born children; the remainder were later-born. In order to permit finer analysis, 16 oldest and 16 only children were included among the first-born subjects, and 16 middle and 16 youngest children were included among the later-born subjects.

Because the age difference between a child and his next oldest and/or next youngest sibling may influence the dependency training he receives, some attempt was made to control this variable. The mean difference between oldest subjects and their next oldest sibling, between middle-born subjects and their next oldest and next youngest siblings, and between youngest subjects and their next youngest sibling was controlled at 3-0 years. Subjects were assigned to control and experimental groups in such a way that (for oldest, middle, and youngest subjects), in addition to control of age differences, a mean family size of 3.5 children was insured within each group.

[1] This study was supported in part by a grant from the National Research Council of Canada (APA-47) to R. H. Walters.

Apparatus

The autokinetic apparatus has been described by Walters and Quinn (1960). With the addition of a set of relays and an electric timer, this apparatus provided a controlled exposure of the stimulus (a point of light 7 feet from the subject) and an automatic recording of the interval between the presentation of the stimulus and the subject's signal that the light had commenced to move (the subject's latency). (During testing, the only illumination in the room was provided by two small night-lights which allowed the experimenter to read the timer and record the subject's responses.)

Procedure

In order to be sure that the subject could identify the stimulus and follow through the instructions correctly, a few preliminary trials were given. The subject was then instructed:

> I am going to turn on the light several times. Each time, after a short pause, the light will begin to move. As soon as you see the light start to move, press this switch (the experimenter indicated a lever to the right of the subject) and hold it down until the light goes out. The light will go out shortly after you press the switch. Then I shall ask you how far you think the light moved that time. Please try and give an answer to the nearest half inch. The light may not move the same distance each time and it may not always move in the same direction. O.K.? The light will come on, you press the switch as soon as you see the light start to move, and then I shall ask you how far you think it moved. O.K.?

Five trials, constituting the presuggestion series, were then given. On each trial the subject's latency and distance estimate were recorded.

Control subjects were now told:

> For the next five trials I'll give you some feedback as to how well you are estimating the distance the light moves. The procedure will be the same as for the first five trials, but after you give each estimate I shall tell you how far the light actually moved that time. O.K.?

For the experimental subjects, instructions were as follows:

> There will be five trials in which an electric shock will be given for any estimate which differs from the correct distance by more than one and a half inches. However, to give you some idea of how well you are doing, I am first going to give you some information as to how close your estimates are. There will be a series of five trials in which I shall tell you how far the light actually moved each time. The procedure will be the same as for the first five trials, but after you give each estimate I shall tell you how far the light moved that time. O.K.?

Five more trials were now given. To each of the subject's estimates the experimenter added a variable amount, averaging 2.5 inches over the five trials. Ex-

perimental and control subjects were treated exactly alike. "Feedback" was given on each trial.

Control subjects were now told:

Now there will be five more trials that will be just like the first five. That is, after you give your estimate I shall give you no information as to how far the light actually moved. O.K.?

In the case of experimental subjects, the experimenter now took a thimble, from which a wire extended to the rear end of the apparatus, and placed it on a finger of the subject's left hand. The subject was now told:

There will be five more trials following the same procedure as the first five. That is, after you give your estimate I'll give you no information as to how far the light actually moved. However, if your estimate differs from the actual distance by more than one and a half inches, you will receive an electric shock through the thimble. The shock is rather severe, but if your judgments are accurate you may not receive any. O.K.?

Five postsuggestion trials were given all subjects. No subject in the experimental group was, in fact, shocked.

Before leaving the experimental room, each subject was handed a pencil and paper and was asked to draw lines 1 inch, 3 inches, and 5 inches in length.

About one month after the experiment all subjects were asked to indicate whether they were definitely willing, probably willing, or definitely not willing to participate in another experiment similar to the one already carried out.

RESULTS

Inspection of data indicated that, at all stages of the experiment, differences between oldest and only subjects and between middle and youngest subjects were negligible. Consequently, results will be given only for the larger groups of first-born and later-born subjects. Statistical analyses are based on the median presuggestion latency and judgment and the median postsuggestion latency and judgment of individual subjects.

Table 1 gives the average of the median latencies of first-born and later-born subjects on the presuggestion trials. The results were in the predicted direction but were significant only at the .06 level.

Before subjects were compared for their responses to suggestion, analyses of their judgments of length, based on their drawings of lines, were carried out. An analysis of variance (Lindquist's, 1953, Type I design), showed that there were no systematic differences among the eight possible groups of subjects (four birth-orders, two conditions). The groups were then combined, and a t test for correlated data was carried out to ascertain whether the judged distance between 1 and 3 inches was the same as the judged distance between 3 and 5 inches. Mean judgments of these two intervals did not differ. On the basis of these analyses, it was decided to treat the judged distances in the autokinetic situation as if these fell on an interval scale that remained relatively constant from group to group. This decision permitted a test of the interaction effect predicted by Hypothesis 2.

TABLE 1

MEANS AND *SD*s (in Seconds) OF PRESUGGESTION LATENCIES OF FIRST-BORN AND LATER-BORN SUBJECTS

(N = 32 in each group)

First-Born		Later-Born	
M	*SD*	*M*	*SD*
9.33	10.61	14.50	15.08

$CR = 1.58$; $p < .06$ (one-tailed test)

Note.—Since heterogeneity of variance was present, a critical ratio was obtained instead of a t ratio.

Since group regression lines were parallel ($F = 1.73$; $p > .05$), an analysis of covariance of subjects' median estimates in the postsuggestion trials was carried out, using median estimates in the presuggestion trials as the covariant. The adjusted group means are given in Table 2; the results of the analysis of covariance are reported in Table 3. It is clear from these tables that first-born subjects and later-born subjects were differentially influenced by the threat of shock. Subsequent t tests showed that the first-born experimental subjects were significantly more suggestible than any of the other three groups, among which there were no significant differences.

Results from the questionnaire are summarized in Table 4. Results for first-born and later-born subjects were first combined to form a 3×2 contingency table. A chi square test showed that subjects who had been threatened with shock showed less willingness to participate in a further experiment than did subjects who had not been so threatened ($p < .01$). The "definitely willing" and "probably willing" categories were then combined for separate tests within the two birth-order groups. First-born subjects who had been threatened with shock were less willing to return than first-born subjects who had not been threatened ($p < .01$), but there was no difference between later-born subjects who had been threatened and those who had not been threatened ($p = .218$ by Fisher's exact probability test [Siegel, 1956]).

DISCUSSION

In the autokinetic situation, first-born subjects responded more quickly than later-born subjects to the suggestion that the light would move. In addition, first-born subjects under the anxiety arousing condition were more suggestible than first-born subjects under the nonanxiety arousing

TABLE 2
ADJUSTED GROUP MEANS (IN INCHES) OF DISTANCE ESTIMATES ON POSTSUGGESTION TRIALS
(N = 16 in each Group)

Experimental		Control	
First-Born	Later-Born	First-Born	Later-Born
5.36	3.57	3.48	4.10

TABLE 3
ANALYSIS OF COVARIANCE OF POSTSUGGESTION DISTANCE ESTIMATES

Source	*SS*	*df*	*MS*	*F*	*p*
Birth Orders	5.269	1	5.269	1.71	> .05
Threat of Shock	7.333	1	7.333	2.38	> .05
Interaction	22.119	1	22.119	7.17	< .01
Within	182.010	59	3.085		

TABLE 4
DISTRIBUTION OF RESPONSES TO QUESTIONNAIRE

Response	First-Born		Later-Born	
	Experimental	Control	Experimental	Control
Definitely willing	2	5	3	7
Probably willing	3	9	7	6
Definitely unwilling	11	2	6	3

condition, whereas later-born subjects seemed to be unaffected by the anxiety condition. These findings give considerable support to Schachter's (1959) theory concerning the effects of birth-order on affiliative behavior.

However, according to this theory, first-born subjects under nonanxiety conditions should be more suggestible than later-born subjects under similar conditions; as Table 2 indicates, there was virtually no difference between the two groups in the postsuggestion trials. This latter finding would be expected on the basis of the Walters-Karal hypothesis (Walters & Karal, 1960) that anxiety is the important motivating factor in determining susceptibility to social influence and is consistent with findings reported by Walters and Ray (1960). Failure to find a significant main effect of anxiety can be attributed to the fact that, judging by their responses to the request to return, later-born subjects were not made anxious by the threat of shock. On the other hand, under the Walters-Karal hypothesis, one would not expect significant differences between first-born and later-born subjects in the presuggestion trials.

Moreover, Staples[2] administered four TAT cards (6GF, 7GF, 8GM, and 10) to 26 first-born and 26 later-born female subjects. Slides were presented in the second half of a regular lecture period, with no previous indication that testing would be carried out. The situation thus roughly corresponded to the "neutral" affiliation arousing conditions of previous studies of affiliative fantasies (Atkinson, Heyns, & Veroff, 1954; Shipley & Veroff, 1952). Standard instructions for the group TAT were used. Stories were scored for affiliative imagery according to the manual provided by Heyns, Veroff, and Atkinson (1958). Scoring was done independently by two scorers, both of whom were ignorant of the birth-order of subjects; discrepancies were subsequently settled by mutual agreement. One point was given for each instance of affiliative imagery; points were summed over the four stories to provide an index of affiliation for each subject. A Mann-Whitney U test indicated that first-born subjects produced more affiliative imagery than later-born subjects (z = 1.64; p = .05, one-tailed test).

According to Schachter (1959), "dependent behavior will be most strongly manifested in conditions of disturbance and anxiety but . . . there should be other non-anxiety related indications of (a relationship between ordinal position and dependence)" (p. 89). Differences between first-born and later-born subjects in the presuggestion trials and on the fantasy measure are indications of this kind.

Walters and Ray (1960) found no difference between first-born and later-born grade-school boys in rate of conditioning to a social reinforcer (approval). This finding contrasts with the birth-order effects found in the present study; it contrasts also with the majority of findings reported by Schachter (1959). However, Schachter's positive findings have been secured mainly when undergraduate girls have served as subjects. In general, the independence training of boys is more severe than that of girls; potential birth-order effects may therefore be obscured in studies of male subjects.

SUMMARY

Sixty-four female undergraduates, 32 first-born and 32 later-born, were given suggestions in the autokinetic situation. Before the postsuggestion trials, half the first-born and half the later-born subjects were told that they would receive shock when their judgments were in error (anxiety condition).

In presuggestion trials first-born subjects responded more rapidly than later-born subjects to the suggestion that the light would move. On the postsuggestion trials first-born subjects under the anxiety condition were more suggestible than first-born subjects under the nonanxiety condition.

[2] Unpublished study carried out in partial fulfilment of the requirements for a graduate course at the University of Toronto.

There was no significant difference between later-born subjects under the anxiety condition and later-born subjects under the nonanxiety condition.

The findings were discussed with reference to hypotheses advanced by Schachter (1959) and Walters and Karal (1960) concerning antecedents of susceptibility to social influence.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

EHRLICH, D. Determinants of verbal commonality and influencibility. Unpublished doctoral thesis, University of Minnesota, 1958.

HAEBERLE, ANN. Interactions of sex, birth order, and dependency with behavior problems and symptoms in emotionally disturbed pre-school children. Paper read at Eastern Psychological Association Annual Meeting, Philadelphia, 1958.

HEYNS, R. W., VEROFF, J., & ATKINSON, J. W. A scoring manual for the affiliation motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 205–218.

LINDQUIST, E. F. *Design and analysis of experiments*. Boston: Houghton Mifflin, 1953.

ROSONOW, C., & WHYTE, ANNE H. The ordinal position of problem children. *Amer. J. Orthopsychiat.*, 1931, **1**, 430–444.

SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford Univer. Press, 1959.

SEARS, R. R. Ordinal position in the family as a psychological variable. *Amer. sociol. Rev.*, 1950, **15**, 397–401.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SLETTO, R. F. Sibling position and juvenile delinquency. *Amer. J. Sociol.*, 1934, **39**, 657–659.

WALTERS, R. H., & KARAL, PEARL. Social deprivation and verbal behavior. *J. Pers.*, 1960, **28**, 89–107.

WALTERS, R. H., & QUINN, M. J. The effects of social and sensory deprivation on autokinetic judgments. *J. Pers.*, 1960, **28**, 210–219.

WALTERS, R. H., & RAY, E. Anxiety, isolation, and reinforcer effectiveness. *J. Pers.*, 1960, **28**, 358–367.

WRIGHTSMAN, L. The effect of small-group membership on level of concern. Unpublished doctoral thesis, University of Minnesota, 1959.

(Received May 31, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 3, 720-722

THE INFLUENCE OF VALENCE ON DISTANCE JUDGMENTS OF HUMAN FACES

JOHN BELOFF AND HALLA BELOFF
Queen's University, Belfast

EARLIER theories of perception, taking their cue from empiricist philosophy, made a sharp distinction between the *descriptive* properties of an object, which included colors and the metric properties of size, shape, and distance, and what one might call the *ascriptive* properties, which included expressive, physiognomic, or affective qualities. The latter, it was held, being invested in the object by dint of association, could be ignored from the strictly perceptual standpoint as belonging rather to the psychology of the emotions. Modern theories of perception no longer allow us to make such a clear cut dichotomy, if only because it is now realized that there is interaction between both types of property. This is true even of inanimate objects but it is especially true of person perception, where the problem is often to ascertain how, to quote Ittelson and Slack (1958), "the feelings and experiences of the observer influence the way in which he perceives the physical characteristics of the person."

These authors have discussed the theoretical issues in this field and reviewed the empirical findings from studies using aniseikonic lenses, the stereoscope, and the Ames Thereness-Thatness apparatus. (The latter is a means by which distance judgments must be made under conditions of minimal cues.)

The present study was generally inspired by this review and specifically by the studies of Smith (1953). Smith, verbally reinforcing the evident expressions on photographed faces, showed that "liked" people with pleasant expressions were placed nearer the observer than "disliked" people with unpleasant expressions, in the Thereness-Thatness apparatus. He used selected photographs from the Frois-Wittmann series. These have the advantage that the impression they make is known, but on the other hand Frois-Wittmann himself posed for all of them, and this must have been embarrassingly obvious to the subjects!

The variable manipulated by Smith, called "emotional loading" by Ittelson and Slack (1958), can perhaps be more incisively thought of as *valence*. Then it becomes clear that the paradigm of a testing situation is indeed one involving distance judgment, in that a stimulus object with positive valence, which "attracts," must be placed too near, and one with negative, "repelling," valence be placed too far. Parenthetically it is interesting to note that in his analysis of adjectives by the Semantic Differential, Osgood (1957) showed "near-far" to have an appreciable loading on the factor of Evaluation, and negligible loadings on the other two factors.

In exploring further in this field we too used the Thereness-Thatness apparatus to test our general proposition that faces hypothesized to have positive valence for our subjects, are set nearer the subjects than faces that have negative valence. Then, aiming to present a situation with stimuli both more natural and more realistic than Smith's (1953), we used principally photographs of people in public life, or personally known to the subjects. Valences were thus not manufactured in the laboratory but were already part of the subjects' general attitude structure.

The following are the several specific examples of the general hypothesis investigated:

1. As the primary test of the hypothesis we tested the influence of valence pure, unconfounded by physiognomic cues, in terms of our "Nonsuch" person. This was a photograph of a nondescript middle-aged man, to whom were ascribed admired and despised roles for alternate subjects. That is, half of them were told that he was a World Health Organization worker in Asia and the other half that he was a United States labor racketeer. If the identical stimulus when made positively valent would be set nearer than when negatively valent, this would be a crucial confirmation.

2. Positive and negative attitudes already anchored to particular faces were investigated by presenting people deemed widely popular in Britain (Prince Philip and President Eisenhower), contrasted with two definitely unpopular (a leader of the Communist Party in Hungary and a member of the Japanese General Staff from the Second World War).[1]

3. Authority figures have been shown by Wittreich and Radcliffe (1956) to be perceived differentially. In our turn we predicted that an authority figure should have negative valence, in comparison with a group of peers, who would be predicted to have approximately neutral valence. Hence a photograph of the Head of the Department was

[1] After some preliminary inquiry we found that among our students at least there were no present day public figures who would be both instantly recognized and abhorred. Hence the negative faces had to be identified by the experimenters, i.e., as Erno Gerö, First Secretary of the Hungarian Communist Party and thus responsible for the causes of the revolution and its brutal suppression, and as a Japanese general executed for atrocities to Allied prisoners of war. We feel confident, however, that this identification did call up existing attitudes.

presented to our subjects (his students) to be compared with a group of six fellow students (those nearest each subject alphabetically by surname).

4. That the self is responded to more favorably in a hidden situation than comparable stimuli, has been shown by the present authors (Beloff & Beloff, 1959). And Ittelson and Slack (1958) report an unpublished study in which it was found that self-photographs were set "differently" from those of a stranger in the Thereness-Thatness apparatus. We predicted more specifically that the self is positively valent, and should be set nearer the subjects than the average of the peer settings. Following Jourard and Secord (1955) we further predicted that this effect would be more striking in the women subjects.

METHOD

Subjects

Twenty-four men and 30 women attending elementary psychology courses at Queen's University served as subjects. They may be considered psychologically naive.

Apparatus

The Thereness-Thatness apparatus is described in detail by Ittelson (1952). We used the simple, mirror model version, the only noteworthy modification being that our table or field was shorter than the one described, being only 6′ in length.

Essentially it consists of two fields of vision perceived as side by side by the subject. The comparison field, containing five comparison posts is viewed binocularly and presents relatively full distance cues. The experimental field consists of only the stimulus figure seemingly suspended in blackness, which is viewed monocularly. Under this condition of reduced cues, the stimulus must be adjusted to be level with the middle comparison post.

Material

The stimulus cards were matt, black and white photographs, in which the variables of size of the face and of the overall dimension (3″ × 4¼″) were kept constant. The Nonsuch face, and those of the public figures were reproduced from clear illustrated magazine photographs. The photographs of the authority figure and of the students themselves were taken by the experimenters.

Procedure

The subject was brought into the completely darkened laboratory and seated at the apparatus. At no point did he see the arrangement of the apparatus. The viewing aperture having been opened, instructions were as follows:

> You see in front of you a photograph of a face beside a series of five rods. We want you to turn the knob which moves the photograph back and forth until you are satisfied that it is level with the middle rod. [Two practice stimuli were adjusted in this way.] We are going to show you a number of such faces. Some will be people you will recognize from public life, others will be people you actually know. This has been done just to add some variety to an otherwise monotonous task. Because your job each time is simply to adjust the photograph until you judge it to be level with the middle post.

Between each stimulus the viewing aperture was closed and the error in distance setting recorded in inches.

All stimuli were presented in a random order, and further randomized with respect to initial distance from the point of equality (5, 10, or 15″ away), and with respect to nearer than or further than, equality.

TABLE 1
MEAN ERROR IN DISTANCE JUDGMENT
(inches)

Stimulus	Combined	Men	Women
Nonsuch +ve	9.08		
Nonsuch −ve	6.86		
Philip	8.62	9.20	8.45
Eisenhower	8.52	7.72	9.44
Gerö	6.86	8.00	6.18
Japanese	6.69	7.40	6.34
Professor	6.93	7.24	6.92
Self	9.14	8.12	10.26
Student mean	8.30	8.36	8.52

Note.—All errors are in the direction of setting the stimulus nearer than the point of equality.

TABLE 2
MEAN DIFFERENCE IN DISTANCE JUDGMENTS

Stimuli	Combined	Men	Women
Nonsuch +ve:Nonsuch −ve	2.22*		
Philip:Gerö	1.76**	1.20*	2.27**
Philip:Japanese	1.93**	1.80**	2.11*
Eisenhower:Gerö	1.66*	−0.28	3.26**
Eisenhower:Japanese	1.83**	0.32	3.10**
Student mean:Professor	1.37**	1.12*	1.60*
Self:Student mean	0.84	−0.24	1.74*

Note.—First stimulus predicted to be nearer than the second. One-tailed tests.
* $p < .05$.
** $p < .01$.

RESULTS AND DISCUSSION

From Table 1 it is seen that the subjects demonstrated consistent and strong constant errors in the direction of placing the stimuli too near themselves. This is of course, irrelevant to our hypothesis. In the apparatus the most important cue to distance is provided by some estimate of the absolute size of the stimulus card. We suspect then that the subjects thought of the photographs as postcard size, while in fact they were considerably smaller.

Also from Table 1 we may construct the order of nearness to the subjects of the whole series of

photographs. Considering both men and women the order is: Nonsuch +ve, Philip, Eisenhower, Peers, Professor, Nonsuch −ve, Gerö, Japanese General.

The specific hypotheses were tested by non-parametric tests for the usual reasons. The Wilcoxon matched-pairs signed-ranks tests was used for all comparisons, except that between the two Nonsuch responses where the Medians test was applied. Table 2 gives the difference in mean distances between the appropriate stimuli and their levels of significance. Let us consider each of our hypotheses in turn.

1. In the case of the nonsuch photograph where the variable of attitude acted without reinforcement of any possible physiognomic factors, the subjects who had been given the favorable label did put the photograph nearer themselves than those who had had the unfavorable one. The difference of 2.22″ is significant at the .05 level. The basic premise has been confirmed.

2. Of the two faces from public life which we predicted would evoke a positive valence only Prince Philip was placed significantly nearer than both the negative figures by both men and women. Eisenhower was distinguished from the negative figures by the women but not by the men. (Is it that they were impervious to his charm and considered only his political record?)[2]

3. Both men and women kept the authority figure "at a distance," that is they set the Professor significantly further back than the average of the fellow students.

4. We predicted that the self-photograph would be set nearer the subjects than the average setting of the peers, and that this phenomenon would be more pronounced among the women. It is seen now that for men the difference in the two settings is indeed negligible, while for women it is significant in the predicted direction at the .05 level.

In general we may conclude that women are influenced to a very significant extent by the valence of the stimuli, even when they were making what seemed to them objective distance comparisons. This would appear to be the case to a lesser extent with men.

Therefore the present study has once again shown that what we perceive is influenced in some degree by our attitude towards the object of perception. What is behind this effect? Because in our experimental arrangement the cues for distance judgment are in fact minimal and perception is thus ambiguous and fluid, we suggest that the variable of assumed size of the stimulus card, which we mentioned before, was being unconsciously overlaid by the process shown most clearly in Smith's (1953) "self-reference" setting. Here the subject is shown only the experimental field of the Thereness-Thatness apparatus and instructed to set the face simply where he would like to see it in relation to himself. This may be thought of as making concrete the metaphor of social distance. In other words, the perceptual responses we obtained were a compromise between an objective size-distance, and a self-reference setting.

[2] From Table 1 it is patent that there are no significant differences between Philip and Eisenhower, or between Gerö and the Japanese general.

SUMMARY

An experiment was conducted to test the hypothesis that photographs of positively valent faces would be set nearer the subject than negative faces, in ostensibly objective distance judgments. The Ames Thereness-Thatness apparatus was used. Several examples of valent faces were compared, e.g., popular vs. unpopular figures from public life, an authority figure vs. peers, and self vs. peers. The responses of women subjects bore out the predictions, those of the men did not do so uniformly.

REFERENCES

BELOFF, HALLA, & BELOFF, J. Unconscious self-evaluation using a stereoscope. *J. abnorm. soc. Psychol.*, 1959, **59**, 275–278.

ITTELSON, W. H. *The Ames demonstrations in perception.* Princeton: Princeton Univer. Press, 1952. P. 25.

ITTELSON, W. H., & SLACK, C. W. The perception of persons as visual objects. In R. Taguiri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. Pp. 210–228.

JOURARD, S. M., & SECORD, P. F. Body-cathexis and personality. *Brit. J. Psychol.*, 1955, **46**, 130–138.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. C. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957. P. 37.

SMITH, G. H. Size-distance judgments of human faces (projected images). *J. gen. Psychol.*, 1953, **49**, 45–64.

WITTREICH, W. J., & RADCLIFFE, K. B. Differences in the perception of an authority figure and a non-authority figure by Navy recruits. *J. abnorm. soc. Psychol.*, 1956, **53**, 383–384.

(Received June 6, 1960)

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition the following special branches are dealt with fully: psychoanalysis, individual and analytical psychology; sexology, criminology, alcoholism, and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry, vocational guidance and mental testing; psychometrics; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1400 pp., including an index classified by author and subject

Subscription price $27.00 per annual volume

EXCERPTA MEDICA FOUNDATION

The New York Academy of Medicine Building

2 East 103 Street **New York 29, N.Y.**

PSYCHOLOGICAL TEST SPECIALISTS

Exclusive publishers and distributors of selected psychological tests

Memory-For-Designs Test (MFD)

Simple drawing test of perceptual-motor coordination, serving as extremely sensitive detector of brain injury of many types. **$8.50 for set of materials**

Group Personality Projective Test (GPPT)

Multiple-choice questions about ambiguous stick-figures make possible rapid and effective measurement of significant, factor-analytically isolated aspects of personality. **$13.50 for set of materials**

Id-Ego-Superego Test (IES)

Subtests give independent estimates of personality components from a modified psychoanalytic point of view. **$28.50 for set of materials**

Kahn Test of Symbol Arrangement (KTSA)

Uses symbolic plastic objects which are arranged and sorted; validly discriminates among normals, neurotics, brain-injured, and psychotics, and indicates accompanying personality dynamics. **$25.00 for set of materials**

Full-Range Picture Vocabulary Test (FRPV)

Brief test of intelligence in two forms using the same 16 plates. Words are defined by choosing pictures. Valuable for intelligence screening, and with the physically handicapped. **$7.50 for set of materials**

Geist Picture Interest Inventory (GPII)

Provides for more valid expression of interests by choices among alternative pictured activities. Both qualitative and quantitative interpretations are possible. **$6.00 for set of materials**

Proverbs Test (PT)

Group and individual forms determine similarity of thought processes to those of various normal and clinical groups. Highly sensitive indicator of psychotic processes. **$4.50 for kit**

Vocational Apperception Test: Advanced Form (VAT:ADV)

Separate series of plates showing male and female college-related occupational activities, accompanied by explanation of use and interpretation as a projective test. **$7.25 for set of materials**

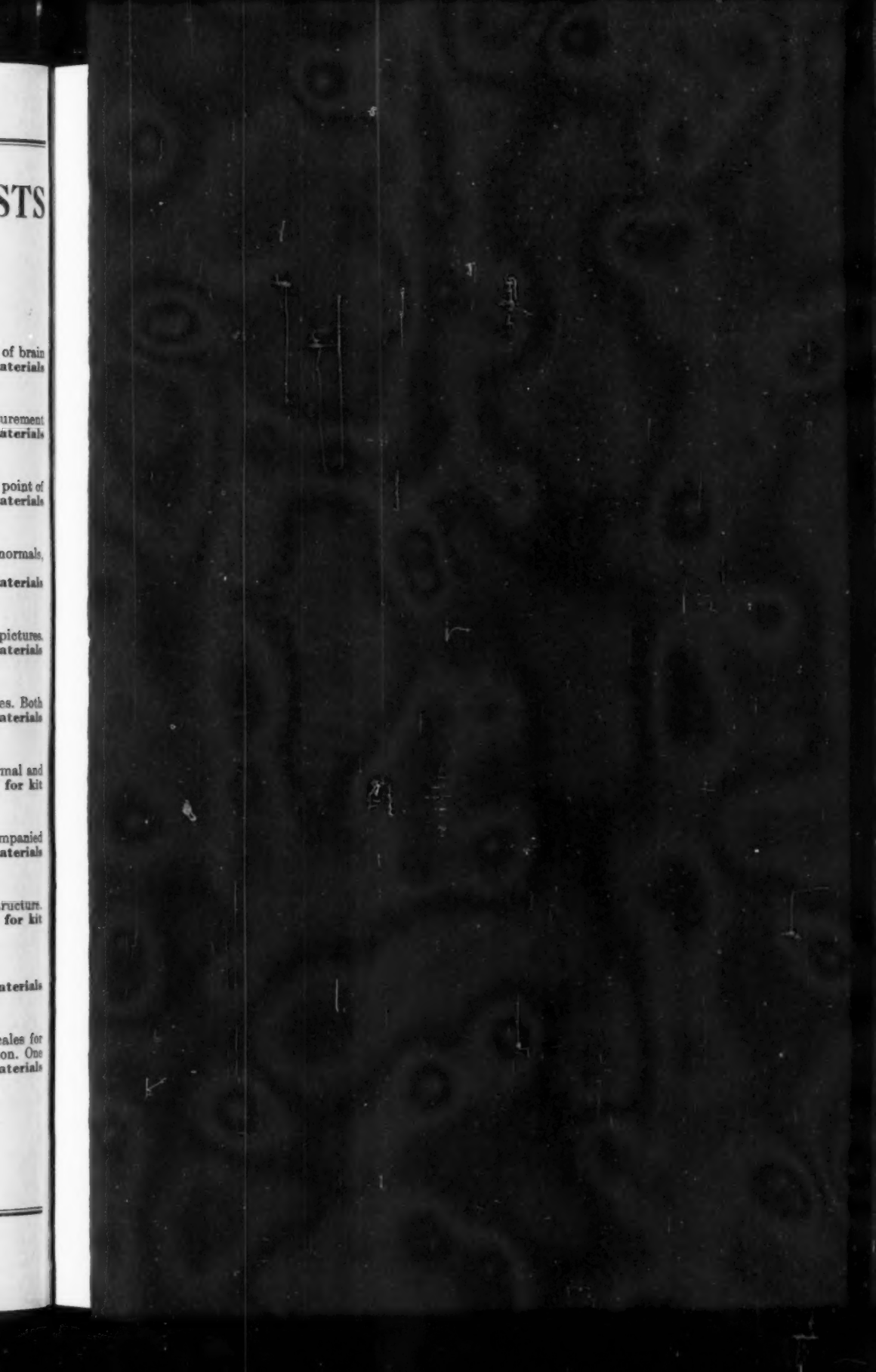
Famous Sayings (FS)

Agreement with proverbs, aphorisms, and folk sayings is analyzed to determine personality structure. Particularly useful in personnel selection. **$4.50 for kit**

It Scale For Children (ITSC)

An experimental picture scale for measuring sex role preference in children. **$15.00 for set of materials**

Kahn Intelligence Tests: Experimental Form (KIT:EXP)

Relatively culture-free intelligence tests requiring no reading or writing. Supplementary scales for assessment of intelligence of blind, ability in concept formation, recall, and motor coordination. One scale can be administered by sign language. **$26.00 for set of materials**

PSYCHOLOGICAL TEST SPECIALISTS
Box 1441 MISSOULA, MONTANA